G-CSF CONJUGATES

COPYRIGHT NOTIFICATION

CROSS-REFERENCES TO RELATED APPLICATIONS

The present application claims priority to and benefit of USSN 60/176,376 filed on January 14, 2000, USSN 60/189,506 filed on March 15, 2000, USSN 60/215,644 filed on June 30, 2000, Danish Application No. PA 2000 00024 filed on January 10, 2000, Danish Application No. PA 2000 00341 filed on March 2, 2000, and Danish Application No. PA 2000 00943 filed on June 16, 2000.

FIELD OF THE INVENTION

The present invention relates to new polypeptides exhibiting granulocyte colony-stimulating factor (G-CSF) activity, to conjugates between a polypeptide exhibiting G-CSF activity and a non-polypeptide moiety, to methods for preparing such polypeptides or conjugates and the use of such polypeptides or conjugates in therapy, in particular for the treatment of leukopenia.

BACKGROUND OF THE INVENTION

The process by which white blood cells grow, divide and differentiate in the bone marrow is called hematopoiesis (Dexter and Spooncer, Ann. Rev. Cell. Biol., 3:423, 1987). Each of the blood cell types arises from pluripotent stem cells. There are generally three classes of blood cells produced *in vivo*: red blood cells (erythrocytes), platelets and white blood cells (leukocytes), the majority of the latter being involved in host immune defence. Proliferation and differentiation of hematopoietic precursor cells are regulated by a family of cytokines, including colony-stimulating factors (CSF's) such as G-CSF and interleukins (Arai *et al.*, Ann. Rev. Biochem., 59:783-836, 1990). The principal biological effect of G-CSF *in vivo* is to stimulate the

growth and development of certain white blood cells known as neutrophilic granulocytes or neutrophils (Welte *et al.*, PNAS-USA 82:1526-1530, 1985, Souza *et al.*, Science, 232:61-65, 1986). When released into the blood stream, neutrophilic granulocytes function to fight bacterial infection.

The amino acid sequence of human G-CSF (hG-CSF) was reported by Nagata *et al.* Nature 319:415-418, 1986. hG-CSF is a monomeric protein that dimerizes the G-CSF receptor by formation of a 2:2 complex of 2 G-CSF molecules and 2 receptors (Horan et al. (1996), *Biochemistry* 35(15): 4886-96). Aritomi *et al.* Nature 401:713-717, 1999 have described the X-ray structure of a complex between hG-CSF and the BN-BC domains of the G-CSF receptor. They identify the following hG-CSF residues as being part of the receptor binding interfaces: G4, P5, A6, S7, S8, L9, P10, Q11, S12, L15, K16, E19, Q20, L108, D109, D112, T115, T116, Q119, E122, E123, and L124. Expression of rhG-CSF in *Escherichia coli*, *Saccharomyces cerevisiae* and mammalian cells has been reported (Souza *et al.*, Science 232:61-65, 1986, Nagata *et al.*, Nature 319: 415-418, 1986, Robinson and Wittrup, Biotechnol. Prog. 11:171-177, 1985).

Recombinant human G-CSF (rhG-CSF) is generally used for treating various forms of leukopenia. Thus, commercial preparations of rhG-CSF are available under the names filgrastim (Gran® and Neupogen®), lenograstim (Neutrogin® and Granocyte®) and nartograstim (Neu-up®). Gran® and Neupogen® are non-glycosylated and produced in recombinant *E. coli* cells. Neutrogin® and Granocyte® are glycosylated and produced in recombinant CHO cells and Neu-up® is non-glycosylated with five amino acids substituted at the N-terminal region of intact rhG-CSF produced in recombinant *E. coli* cells.

A few protein-engineered variants of hG-CSF have been reported (US 5,581,476, US 5,214,132, US 5,362,853, US 4,904,584 and Riedhaar-Olson *et al.* Biochemistry 35: 9034-9041, 1996). Modification of hG-CSF and other polypeptides so as to introduce at least one additional carbohydrate chain as compared to the native polypeptide has been suggested (US 5,218,092). It is stated that the amino acid sequence of the polypeptide may be modified by amino acid substitution, amino acid deletion or amino acid insertion so as to effect addition of an additional carbohydrate chain. In addition, polymer modifications of native hG-CSF, including attachment of PEG groups, have been reported (Satake-Ishikawa *et al.*, Cell Structure and

Function 17:157-160, 1992, US 5,824,778, US 5,824,784, WO 96/11953, WO 95/21629, WO 94/20069).

Bowen et al., Experimental Hematology 27 (1999), 425-432 disclose a study of the relationship between molecule mass and duration of activity of PEG-conjugated G-CSF mutein. An apparent inverse correlation was suggested between molecular weight of the PEG moieties conjugated to the protein and *in vitro* activity, whereas *in vivo* activities increased with increasing molecular weight. It is speculated that a lower affinity of the conjugates act to increase the half-life, because receptor-mediated endocytosis is an important mechanism regulating levels of hematopoietic growth factors.

The commercially available rhG-CSF has a short-term pharmacological effect and must often be administered more once a day for the duration of the leukopenic state. A molecule with a longer circulation half-life would decrease the number of administrations necessary to alleviate the leukopenia and prevent consequent infections. Another problem with currently available rG-CSF products is the occurrence of dose-dependent bone pain. Since bone pain is experienced by patients as a significant side effect of treatment with rG-CSF, it would be desirable to provide a rG-CSF product that does not cause bone pain, either by means of a product that inherently does not have this effect or that is effective in a sufficiently small dose that no bone pain is caused. Thus, there is clearly a need for improved recombinant G-CSF-like molecules.

With respect to the half-life, one way to increase the circulation half-life of a protein is to ensure that clearance of the protein, in particular via renal clearance and receptor-mediated clearance, is reduced. This may be achieved by conjugating the protein to a chemical moiety which is capable of increasing the apparent size, thereby reducing renal clearance and increasing the *in vivo* half-life. Furthermore, attachment of a chemical moiety to the protein may effectively block proteolytic enzymes from physical contact with the protein, thus preventing degradation by non-specific proteolysis. Polyethylene glycol (PEG) is one such chemical moiety that has been used in the preparation of therapeutic protein products. A G-CSF molecule modified with a single, N-terminally linked PEG group, termed SD/01, is currently undergoing clinical trial. Although this PEGylated G-CSF molecule has been shown to have an increased half-life compared to non-PEGylated G-CSF, it has a low activity and must therefore be administered in rather high doses. As a result, the problem of bone pain discussed above has not

been solved by this molecule. Furthermore, there is still substantial room for improvement of the known G-CSF molecules in terms of half-life.

A need therefore still exists for providing novel molecules exhibiting G-CSF activity that are useful in the treatment of leukopenia, and which have are improved in terms of an increased half-life and a reduction in side effects. The present invention relates to such molecules.

SUMMARY OF THE INVENTION

More specifically, the present invention relates to specific conjugates comprising a polypeptide exhibiting G-CSF activity and a non-polypeptide moiety, methods for their preparation and their use in medical treatment and in the preparation of pharmaceuticals. Accordingly, in a first aspect the invention relates to various specific conjugates comprising a polypeptide exhibiting G-CSF activity and having an amino acid sequence that differs from the known amino acid sequence of human G-CSF as shown in SEQ ID NO:1 in at least one specified altered amino acid residue comprising an attachment group for a non-polypeptide moiety, and having at least one non-polypeptide moiety attached to an attachment group of the polypeptide. The conjugate of the present invention has one or more improved properties as compared to commercially available rhG-CSF, including increased functional *in vivo* half-life, increased serum half-life, reduced side effects, reduced immunogenicity and/or increased bioavailability. Consequently, medical treatment with a conjugate of the invention offers a number of advantages over the currently available G-CSF compounds.

In a further aspect the invention relates to polypeptides exhibiting G-CSF activity and which form part of a conjugate of the invention. The polypeptides of the invention are contemplated to be useful as such for therapeutic, diagnostic or other purposes, but find particular interest as intermediate products for the preparation of a conjugate of the invention.

In a further aspect the invention relates to a polypeptide conjugate comprising a polypeptide exhibiting G-CSF activity, which comprises an amino acid sequence that differs from the amino acid sequence of hG-CSF (with the amino acid sequence shown in SEQ ID NO:1) in at least one amino acid residue selected from an introduced or removed amino acid residue comprising an attachment group for a non-polypeptide moiety, and a sufficient number or type of non-polypeptide moieties to provide the conjugate with an increased half-life compared to known recombinant G-CSF products.

In still further aspects the invention relates to methods for preparing a conjugate of the invention, including nucleotide sequences encoding a polypeptide of the invention, expression vectors comprising such a nucleotide sequence, and host cells comprising such a nucleotide sequence or expression vector.

In final aspects the invention relates to a composition comprising a conjugate or polypeptide of the invention, a method for preparing a pharmaceutical composition, use of a conjugate or composition of the invention as a pharmaceutical, and a method of treating a mammal with such composition. In particular, the polypeptide, conjugate or composition of the invention may be used to prevent infection in cancer patients undergoing certain types of radiation therapy, chemotherapy, and bone marrow transplantations, to mobilize progenitor cells for collection in peripheral blood progenitor cell transplantations, for treatment of severe chronic or relative leukopenia, irrespective of cause, and to support treatment of patients with acute myeloid leukaemia. Additionally, the polypeptide, conjugate or composition of the invention may be used for treatment of AIDS or other immunodeficiency diseases as well as bacterial infections.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: The *in vivo* half-lives of rhG-CSF (Neupogen®) and SPA-PEG 5000-conjugated hG-CSF K16R K34R K40R Q70K Q90K Q120K

Figure 2: The *in vivo* biological activities of rhG-CSF (Neupogen®), SPA-PEG 5000-conjugated hG-CSF K16R K34R K40R Q70K Q120K and SPA-PEG 5000-conjugated hG-CSF K16R K34R K40R Q70K Q90K Q120K.

Figure 3: The *in vivo* biological activities of rhG-CSF (Neupogen®), SPA-PEG 12000-conjugated hG-CSF K16R K34R K40R and different doses of SPA-PEG 5000-conjugated hG-CSF K16R K34R K40R Q70K Q90K Q120K.

DETAILED DISCUSSION

DEFINITIONS

In the context of the present application and invention the following definitions apply:

The term "conjugate" is intended to indicate a heterogeneous molecule formed by the covalent attachment of one or more polypeptides, typically a single polypeptide, to one or more non-polypeptide moieties such as polymer molecules, lipophilic compounds, carbohydrate moieties or organic derivatizing agents. The term covalent attachment means that the polypeptide and the non-polypeptide moiety are either directly covalently joined to one another, or else are indirectly covalently joined to one another through an intervening moiety or moieties, such as a bridge, spacer, or linkage moiety or moieties. Preferably, the conjugate is soluble at relevant concentrations and conditions, i.e., soluble in physiological fluids such as blood. The term "non-conjugated polypeptide" may be used about the polypeptide part of the conjugate.

The term "polypeptide" may be used interchangeably herein with the term "protein".

The "polymer molecule" is a molecule formed by covalent linkage of two or more monomers, wherein none of the monomers is an amino acid residue, except where the polymer is human albumin or another abundant plasma protein. The term "polymer" may be used interchangeably with the term "polymer molecule". The term is intended to cover carbohydrate molecules, although, normally, the term is not intended to cover the type of carbohydrate molecule which is attached to the polypeptide by *in vivo* N- or O-glycosylation (as further described below), since such molecule is referred to herein as "an oligosaccharide moiety". Except where the number of polymer molecule(s) is expressly indicated every reference to "a polymer", "a polymer molecule", "the polymer" or "the polymer molecule" contained in a polypeptide of the invention or otherwise used in the present invention shall be a reference to one or more polymer molecule(s).

The term "attachment group" is intended to indicate an amino acid residue group of the polypeptide capable of coupling to the relevant non-polypeptide moiety. For instance, for polymer conjugation, in particular to PEG, a frequently used attachment group is the ε-amino group of lysine or the N-terminal amino group. Other polymer attachment groups include a free carboxylic acid group (e.g.,, that of the C-terminal amino acid residue or of an aspartic acid or

glutamic acid residue), suitably activated carbonyl groups, oxidized carbohydrate moieties and mercapto groups. Useful attachment groups and their matching non-peptide moieties are apparent from the table below.

Attachment group	Amino acid	Examples of non-peptide moiety	Conjugation method/-Activated PEG	Reference
$-NH_2$	N-terminal, Lys, His, Arg	Polymer, e.g., PEG, with amide or imine group	mPEG-SPA Tresylated mPEG	Shearwater Inc. Delgado et al., critical reviews in Therapeutic Drug Carrier Systems 9(3,4):249-304 (1992)
-COOH	C-term, Asp, Glu	Polymer, e.g., PEG, with ester or amide group Oligosaccharide moiety	mPEG-Hz *In vitro* coupling	Shearwater Inc.
-SH	Cys	Polymer, e.g., PEG, with disulfide, maleimide or vinyl sulfone group Oligosaccharide moiety	PEG-vinylsulphone PEG-maleimide *In vitro* coupling	Shearwater Inc. Delgado et al., critical reviews in Therapeutic Drug Carrier Systems 9(3,4):249-304 (1992)
-OH	Ser, Thr, -OH, Lys	Oligosaccharide moiety PEG with ester, ether, carbamate, carbonate	*In vivo* O-linked glycosylation	
$-CONH_2$	Asn as part of an N-glycosylation site	Oligosaccharide moiety Polymer, e.g., PEG	*In vivo N*-glycosylation	
Aromatic residue	Phe, Tyr, Trp	Oligosaccharide moiety	*In vitro* coupling	
$-CONH_2$	Gln	Oligosaccharide moiety	*In vitro* coupling	Yan and Wold, Biochemistry, 1984, Jul 31; 23(16): 3759-65
Aldehyde Ketone	Oxidized oligo-saccharide	Polymer, e.g., PEG, PEG-hydrazide	PEGylation	Andresz et al., 1978, Makromol. Chem. 179:301, WO 92/16555, WO 00/23114

Guanidino	Arg	Oligosaccharide moiety	*In vitro* coupling	Lundblad and Noyes, Chemical Reagents for Protein Modification, CRC Press Inc., Florida, USA
Imidazole ring	His	Oligosaccharide moiety	*In vitro* coupling	As for guanidine

For *in vivo* N-glycosylation, the term "attachment group" is used in an unconventional way to indicate the amino acid residues constituting an N-glycosylation site (with the sequence N-X'-S/T/C-X'', wherein X' is any amino acid residue except proline, X'' any amino acid residue which may or may not be identical to X' and which preferably is different from proline, N is asparagine, and S/T/C is either serine, threonine or cysteine, preferably serine or threonine, and most preferably threonine). Although the asparagine residue of the N-glycosylation site is where the oligosaccharide moiety is attached during glycosylation, such attachment cannot be achieved unless the other amino acid residues of the N-glycosylation site are present. Accordingly, when the non-peptide moiety is an oligosaccharide moiety and the conjugation is to be achieved by N-glycosylation, the term "amino acid residue comprising an attachment group for the non-peptide moiety" as used in connection with alterations of the amino acid sequence of the polypeptide of interest is to be understood as meaning that one or more amino acid residues constituting an N-glycosylation site are to be altered in such a manner that either a functional N-glycosylation site is introduced into the amino acid sequence or removed from said sequence.

In the present application, amino acid names and atom names (e.g., CA, CB, NZ, N, O, C, etc.) are used as defined by the Protein DataBank (PDB) (www.pdb.org), which is based on the IUPAC nomenclature (IUPAC Nomenclature and Symbolism for Amino Acids and Peptides (residue names, atom names etc.), *Eur. J. Biochem.*, 138, 9-37 (1984) together with their corrections in *Eur. J. Biochem.*, 152, 1 (1985). The term "amino acid residue" is primarily intended to indicate an amino acid residue contained in the group consisting of the 20 naturally occurring amino acids, i.e.,, alanine (Ala or A), cysteine (Cys or C), aspartic acid (Asp or D), glutamic acid (Glu or E), phenylalanine (Phe or F), glycine (Gly or G), histidine (His or H), isoleucine (Ile or I), lysine (Lys or K), leucine (Leu or L), methionine (Met or M), asparagine (Asn or N), proline (Pro or P), glutamine (Gln or Q), arginine (Arg or R), serine (Ser or S), threonine (Thr or T), valine (Val or V), tryptophan (Trp or W), and tyrosine (Tyr or Y) residues.

The terminology used for identifying amino acid positions/substitutions is illustrated as follows: F13 indicates position number 13 occupied by a phenylalanine residue in the reference amino acid sequence. F13K indicates that the phenylalanine residue of position 13 has been substituted with a lysine residue. Unless otherwise indicated, the numbering of amino acid residues made herein is made relative to the amino acid sequence of hG-CSF shown in SEQ ID NO:1. Alternative substitutions are indicated with a "/", e.g., Q67D/E means an amino acid sequence in which glutamine in position 67 is substituted with either aspartic acid or glutamic acid. Multiple substitutions are indicated with a "+", e.g., S53N+G55S/T means an amino acid sequence which comprises a substitution of the serine residue in position 53 with an asparagine residue and a substitution of the glycine residue in position 55 with a serine or a threonine residue.

The term "nucleotide sequence" is intended to indicate a consecutive stretch of two or more nucleotide molecules. The nucleotide sequence may be of genomic, cDNA, RNA, semisynthetic or synthetic origin, or any combination thereof.

The term "polymerase chain reaction" or "PCR" generally refers to a method for amplification of a desired nucleotide sequence *in vitro*, as described, for example, in US 4,683,195. In general, the PCR method involves repeated cycles of primer extension synthesis, using oligonucleotide primers capable of hybridising preferentially to a template nucleic acid.

"Cell", "host cell", "cell line" and "cell culture" are used interchangeably herein and all such terms should be understood to include progeny resulting from growth or culturing of a cell. "Transformation" and "transfection" are used interchangeably to refer to the process of introducing DNA into a cell.

"Operably linked" refers to the covalent joining of two or more nucleotide sequences, by means of enzymatic ligation or otherwise, in a configuration relative to one another such that the normal function of the sequences can be performed. For example, the nucleotide sequence encoding a presequence or secretory leader is operably linked to a nucleotide sequence for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide: a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the nucleotide sequences being linked are contiguous and, in the case of a secretory leader,

contiguous and in reading phase. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, then synthetic oligonucleotide adaptors or linkers are used, in conjunction with standard recombinant DNA methods.

The term "introduce" refers to introduction of an amino acid residue comprising an attachment group for a non-polypeptide moiety, in particular by substitution of an existing amino acid residue, or alternatively by insertion of an additional amino acid residue. The term "remove" refers to removal of an amino acid residue comprising an attachment group for a non-polypeptide moiety, in particular by substitution of the amino acid residue to be removed by another amino acid residue, or alternatively by deletion (without substitution) of the amino acid residue to be removed.

When substitutions are performed in relation to a parent polypeptide, they are preferably "conservative substitutions," in other words substitutions performed within groups of amino acids with similar characteristics, e.g., small amino acids, acidic amino acids, polar amino acids, basic amino acids, hydrophobic amino acids and aromatic amino acids.

Preferred substitutions in the present invention may in particular be chosen from among the conservative substitution groups listed in the table below.

Conservative substitution groups:

1	Alanine (A)	Glycine (G)	Serine (S)	Threonine (T)
2	Aspartic acid (D)	Glutamic acid (E)		
3	Asparagine (N)	Glutamine (Q)		
4	Arginine (R)	Histidine (H)	Lysine (K)	
5	Isoleucine (I)	Leucine (L)	Methionine (M)	Valine (V)
6	Phenylalanine (F)	Tyrosine (Y)	Tryptophan (W)	

The term "immunogenicity" as used in connection with a given substance is intended to indicate the ability of the substance to induce a response from the immune system. The immune response may be a cell or antibody mediated response (see, e.g.,, Roitt: Essential Immunology (8th Edition, Blackwell) for further definition of immunogenicity). Normally, reduced antibody reactivity will be an indication of reduced immunogenicity. The reduced

immunogenicity may be determined by use of any suitable method known in the art, e.g., *in vivo* or *in vitro*.

The term "functional *in vivo* half-life" is used in its normal meaning, i.e.,, the time at which 50% of the biological activity of the polypeptide or conjugate is still present in the body/target organ, or the time at which the activity of the polypeptide or conjugate is 50% of the initial value. As an alternative to determining functional *in vivo* half-life, "serum half-life" may be determined, i.e.,,the time in which 50% of the polypeptide or conjugate molecules circulate in the plasma or bloodstream prior to being cleared. Alternative terms to serum half-life include "plasma half-life", "circulating half-life", "serum clearance", "plasma clearance" and "clearance half-life". The polypeptide or conjugate is cleared by the action of one or more of the reticuloendothelial systems (RES), kidney, spleen or liver, by receptor-mediated degradation, or by specific or non-specific proteolysis, in particular by the action of receptor-mediated clearance and renal clearance. Normally, clearance depends on size (relative to the cutoff for glomerular filtration), charge, attached carbohydrate chains, and the presence of cellular receptors for the protein. The functionality to be retained is normally selected from proliferative or receptor-binding activity. The functional *in vivo* half-life and the serum half-life may be determined by any suitable method known in the art as further discussed in the Materials and Methods section below.

The term "increased" as used about the functional *in vivo* half-life or serum half-life is used to indicate that the relevant half-life of the conjugate or polypeptide is statistically significantly increased relative to that of a reference molecule, such as a non-conjugated hG-CSF (e.g., Neupogen®) as determined under comparable conditions. For instance, the relevant half-life may increased by at least about 25%, such as by at least about 50%, e.g., by at least about 100%, 200%, 500% or 1000%.

The term "renal clearance" is used in its normal meaning to indicate any clearance taking place by the kidneys, e.g., by glomerular filtration, tubular excretion or tubular elimination. Renal clearance depends on physical characteristics of the conjugate, including size (diameter), symmetry, shape/rigidity and charge. Reduced renal clearance may be established by any suitable assay, e.g., an established *in vivo* assay. Typically, renal clearance is determined by administering a labelled (e.g., radioactive or fluorescent labelled) polypeptide conjugate to a patient and measuring the label activity in urine collected from the patient. Reduced renal

clearance is determined relative to a corresponding reference polypeptide, e.g., the corresponding non-conjugated polypeptide, a non-conjugated corresponding wild-type polypeptide or another conjugated polypeptide (such as a conjugated polypeptide not according to the invention), under comparable conditions. Preferably, the renal clearance rate of the conjugate is reduced by at least 50%, preferably by at least 75%, and most preferably by at least 90% compared to a relevant reference polypeptide.

Generally, activation of the receptor is coupled to receptor-mediated clearance (RMC) such that binding of a polypeptide to its receptor without activation does not lead to RMC, while activation of the receptor leads to RMC. The clearance is due to internalisation of the receptor-bound polypeptide with subsequent lysosomal degradation. Reduced RMC may be achieved by designing the conjugate so as to be able to bind and activate a sufficient number of receptors to obtain optimal *in vivo* biological response and avoid activation of more receptors than required for obtaining such response. This may be reflected in reduced *in vitro* bioactivity and/or increased off-rate.

Typically, reduced *in vitro* bioactivity reflects reduced efficacy/efficiency and/or reduced potency and may be determined by any suitable method for determining any of these properties. For instance, *in vitro* bioactivity may be determined in a luciferase based assay (see Materials and Methods). The *in vitro* bioactivity of the conjugate may be reduced by at least 30%, such as by at least 50%, 60% or 75%, e.g., by at least 90%, as compared to the *in vitro* bioactivity of a corresponding reference polypeptide, determined under comparable conditions. Expressed differently, the conjugate may have an *in vitro* bioactivity that is as small as about 1%, typically at least about 2%, such as at least about 5%, of that of the corresponding non-conjugated polypeptide or of the corresponding wild-type polypeptide. For instance, the *in vitro* bioactivity may be in the range of about 1-50% of that of the reference polypeptide, such as about 2-40%, e.g., about 5-25%, determined under comparable conditions. In cases where reduced *in vitro* bioactivity is desired in order to reduce receptor-mediated clearance, it will be clear that sufficient bioactivity to obtain the desired receptor activation must be nevertheless be maintained.

Preferably, the increased off-rate between the polypeptide conjugate and its receptor is of a magnitude resulting in the polypeptide conjugate being released from its receptor before any substantial internalisation of the receptor-ligand complex has taken place. The

receptor-polypeptide binding affinity (including off-rate) may be determined as described in the Materials and Methods section herein. The *in vitro* RMC may be determined by labelling (e.g., radioactive or fluorescent labelling) the polypeptide conjugate, stimulating cells comprising the receptor for the polypeptide, washing the cells, and measuring label activity. Alternatively, the conjugate may be exposed to cells expressing the relevant receptor. After an appropriate incubation time the supernatant is removed and transferred to a well containing similar cells. The biological response of these cells to the supernatant is determined relative to a non-conjugated polypeptide or another reference polypeptide, and this is a measure of the extent of the reduced RMC.

Normally, reduced *in vitro* bioactivity of the conjugate is obtained as a consequence of its modification by a non-polypeptide moiety. However, in order to further reduce *in vitro* bioactivity or for other reasons it may be of interest to modify the polypeptide part of the conjugate further. For instance, in one embodiment at least one amino acid residue located at or near a receptor binding site of the polypeptide may be substituted with another amino acid residue as compared to the corresponding wild-type polypeptide so as to obtain reduced *in vitro* bioactivity. The amino acid residue to be introduced by substitution may be any amino acid residue capable of reducing *in vitro* bioactivity of the conjugate. Conveniently, the introduced amino acid residue comprises an attachment group for the non-polypeptide moiety as defined herein. In particular, when the non-polypeptide moiety is a polymer molecule such as PEG, the amino acid residue to be introduced may be a lysine residue.

The term "exhibiting G-CSF activity" is intended to indicate that the polypeptide or conjugate has one or more of the functions of native G-CSF, in particular hG-CSF with the amino acid sequence shown in SEQ ID NO:1, including the capability to bind to a G-CSF receptor (Fukunaga *et al.*, J. Bio. Chem, 265:14008, 1990). The G-CSF activity is conveniently assayed using the primary assay described in the Materials and Methods section hereinafter. The polypeptide "exhibiting" G-CSF activity is considered to have such activity when it displays a measurable function, e.g., a measurable proliferative activity or a receptor binding activity (e.g., as determined by the primary assay described in the Materials and Methods section). The polypeptide exhibiting G-CSF activity may also be termed " G-CSF molecule" herein.

The term "parent G-CSF" or "parent polypeptide" is intended to indicate the molecule to be modified in accordance with the present invention. The parent G-CSF is normally

hG-CSF or a variant thereof. A "variant" is a polypeptide which differs in one or more amino acid residues from a parent polypeptide, normally in 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 amino acid residues. Examples of rhG-CSF include filgrastim (Gran® and Neupogen®), lenograstim (Neutrogin® and Granocyte®) and nartograstim (Neu-up®).

Conjugate of the invention

As stated above, in a first aspect the invention relates to a conjugate comprising a polypeptide exhibiting G-CSF activity, which comprises an amino acid sequence that differs from the amino acid sequence of SEQ ID NO:1 in at least one amino acid residue selected from specified introduced or removed amino acid residues comprising an attachment group for a non-polypeptide moiety, and at least one non-polypeptide moiety attached to an attachment group of the polypeptide. The amino acid residues to be introduced and/or removed are described in further detail in the following sections. It will be understood that the conjugate itself also exhibits G-CSF activity.

By removing and/or introducing an amino acid residue comprising an attachment group for the non-polypeptide moiety it is possible to specifically adapt the polypeptide so as to make the molecule more susceptible to conjugation to the non-polypeptide moiety of choice, to optimize the conjugation pattern (e.g., to ensure an optimal distribution of non-polypeptide moieties on the surface of the G-CSF molecule and to ensure that only the attachment groups intended to be conjugated are present in the molecule) and thereby obtain a new conjugate molecule which has G-CSF activity and in addition one or more improved properties as compared to G-CSF molecules available today.

The polypeptide can be of any origin, e.g., of mammalian origin. In some embodiments, the polypeptide is of human origin.

In preferred embodiments of the present invention more than one amino acid residue of the polypeptide with G-CSF activity is altered, e.g., the alteration embraces removal as well as introduction of amino acid residues comprising an attachment group for the non-polypeptide moiety of choice.

In addition to the amino acid alterations disclosed herein aimed at removing and/or introducing attachment sites for the non-polypeptide moiety, it will be understood that the amino acid sequence of the polypeptide of the invention can, if desired, contain other alterations that need not be related to introduction or removal of attachment sites, i.e., other substitutions,

insertions or deletions. These may, for example, include truncation of the N- and/or C-terminus by one or more amino acid residues, or addition of one or more extra residues at the N- and/or C-terminus, e.g., addition of a methionine residue at the N-terminus.

The conjugate of the invention has one or more of the following improved properties as compared to hG-CSF, in particular as compared to rhG-CSF (e.g., filgrastim, lenograstim or nartograstim) or known hG-CSF variants: increased functional *in vivo* half-life, increased serum half-life, reduced renal clearance, reduced receptor-mediated clearance, reduced side effects such as bone pain, and reduced immunogenicity.

It will be understood that the amino acid residue comprising an attachment group for a non-polypeptide moiety, whether it be removed or introduced, will be selected on the basis of the nature of the non-polypeptide moiety of choice and, in most instances, on the basis of the method by which conjugation between the polypeptide and the non-polypeptide moiety is to be achieved. For instance, when the non-polypeptide moiety is a polymer molecule such as a polyethylene glycol or polyalkylene oxide derived molecule amino acid residues comprising an attachment group may be selected from the group consisting of lysine, cysteine, aspartic acid, glutamic acid, histidine and arginine. When conjugation to a lysine residue is to be achieved, a suitable activated molecule is e.g., mPEG-SPA from Shearwater Polymers, Inc., oxycarbonyl-oxy-N-dicarboxyimide-PEG (US 5,122,614), or PEG available from PolyMASC Pharmaceuticals plc. The first of these will be illustrated further below.

In order to avoid too much disruption of the structure and function of the parent hG-CSF molecule, the total number of amino acid residues to be altered in accordance with the present invention, e.g., as described in the subsequent sections herein, (as compared to the amino acid sequence shown in SEQ ID NO:1) will typically not exceed 15. The exact number of amino acid residues and the type of amino acid residues to be introduced or removed depends in particular on the desired nature and degree of conjugation (e.g., the identity of the non-polypeptide moiety, how many non-polypeptide moieties it is desirable or possible to conjugate to the polypeptide, where conjugation is desired or should be avoided, etc.). Preferably, the polypeptide part of the conjugate of the invention or the polypeptide of the invention comprises an amino acid sequence which differs in 1-15 amino acid residues from the amino acid sequence shown in SEQ ID NO:1, typically in 2-10 amino acid residues, e.g., in 3-8 amino acid residues, such as 4-6 amino acid residues, from the amino acid sequence shown in SEQ ID NO:1. Thus,

normally the polypeptide part of the conjugate or the polypeptide of the invention comprises an amino acid sequence which differs from the amino acid sequence shown in SEQ ID NO:1 in 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 amino acid residues.

The polypeptide part of the conjugate will typically have an amino acid sequence with at least about 80% identity with SEQ ID NO:1, preferably at least about 90%, such as at least about 95%. Amino acid sequence homology/identity is conveniently determined from aligned sequences, using e.g., the ClustalW program, version 1.8, June 1999, using default parameters (Thompson et al., 1994, ClustalW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Research, 22: 4673-4680) or from the PFAM families database version 4.0 (http://pfam.wustl.edu/) (*Nucleic Acids Res.* 1999 Jan 1; 27(1):260-2) by use of GENEDOC version 2.5 (Nicholas, K.B., Nicholas H.B. Jr., and Deerfield, D.W. II. 1997 GeneDoc: Analysis and Visualization of Genetic Variation, EMBNEW.NEWS 4:14; Nicholas, K.B. and Nicholas H.B. Jr. 1997 GeneDoc: Analysis and Visualization of Genetic Variation).

In a preferred embodiment one difference between the amino acid sequence of the polypeptide and the amino acid sequence shown in SEQ ID NO:1 is that at least one and often more, e.g., 1-15, amino acid residues comprising an attachment group for the non-polypeptide moiety has been introduced, preferably by substitution, into the amino acid sequence. Thereby, the polypeptide part is altered in the content of the specific amino acid residues to which the non-polypeptide moiety of choice binds, whereby a more efficient, specific and/or extensive conjugation is achieved. For instance, when the total number of amino acid residues comprising an attachment group for the non-polypeptide of choice is altered to an optimised level, the clearance of the conjugate is typically significantly reduced, due to the altered shape, size and/or charge of the molecule achieved by the conjugation. Furthermore, when the total number of amino acid residues comprising an attachment group for the non-polypeptide of choice is increased, a greater proportion of the polypeptide molecule is shielded by the non-polypeptide moieties of choice, leading to a lower immune response.

The term "one difference" as used in the present application is intended to allow for additional differences being present. Accordingly, in addition to the specified amino acid difference, other amino acid residues than those specified may be mutated.

In a further preferred embodiment one difference between the amino acid sequence of the polypeptide and the amino acid sequence shown in SEQ ID NO:1 is that at least one and preferably more, e.g., 1-15, amino acid residues comprising an attachment group for the non-polypeptide moiety has/have been removed, preferably by substitution, from the amino acid sequence. By removing one or more amino acid residues comprising an attachment group for the non-polypeptide moiety of choice it is possible to avoid conjugation to the non-polypeptide moiety in parts of the polypeptide in which such conjugation is disadvantageous, e.g., in amino acid residues located at or near a functional site of the polypeptide (since conjugation at such a site may result in inactivation or reduced G-CSF activity of the resulting conjugate due to impaired receptor recognition). In the present context the term "functional site" is intended to indicate one or more amino acid residues which is/are essential for or otherwise involved in the function or performance of hG-CSF. Such amino acid residues are a part of the functional site. The functional site may be determined by methods known in the art and is preferably identified by analysis of a structure of the polypeptide complexed to a relevant receptor, such as the hG-CSF receptor (See Aritomi *et al.* Nature 401:713-717, 1999).

In a still further preferred embodiment, the amino acid sequence of the polypeptide differs from the amino acid sequence shown in SEQ ID NO:1 in that a) at least one specified amino acid residue comprising an attachment group for the non-polypeptide moiety and present in the amino acid sequence shown in SEQ ID NO:1 has been removed, preferably by substitution, and b) at least one specified amino acid residue comprising an attachment group for the non-polypeptide moiety has been introduced into the amino acid sequence, preferably by substitution, the specified amino acid residues being any of those described in the subsequent sections herein. This embodiment is considered of particular interest in that it is possible to specifically design the polypeptide so as to obtain an optimal conjugation to the non-polypeptide moiety of choice. For instance, by introducing and removing selected amino acid residues as disclosed in the following sections it is possible to ensure an optimal distribution of attachment groups for the non-polypeptide moiety of choice, which gives rise to a conjugate in which the non-polypeptide moieties are placed so as to a) effectively shield epitopes and other surface parts of the polypeptide and b) ensure an optimal Stokes radius of the conjugate, without causing too much structural disruption and thereby impair the function of the polypeptide.

The conjugate of the invention will in general comprise a sufficient number and type of non-polypeptide moieties to provide the conjugate with an increased functional *in vivo* half-life and/or serum half-life as compared to hG-CSF, e.g., filgrastim, lenograstim or nartograstim, and preferably as compared to rhG-CSF comprising a single N-terminally attached 20 kDa PEG moiety. The increased functional *in vivo* half-life is conveniently determined as described in the Materials and Methods section herein.

The conjugate of the invention may comprise at least one non-conjugated, conjugatable attachment group for the non-polypeptide moiety. In the present context the term "conjugatable attachment group" is intended to indicate an attachment group that is located in a position of the polypeptide where it is accessible for conjugation, and that but for special precautions is conjugated to the relevant non-polypeptide moiety when subjected to conjugation. For instance, such attachment group may be part of an amino acid residue involved in or otherwise essential for the polypeptide to exert its activity. A convenient way to avoid conjugation of an otherwise conjugatable attachment group is to shield the attachment group by means of a helper molecule, e.g., as described in the section entitled "Blocking of the functional site". It will be understood that the number of non-conjugated, conjugatable attachment groups depends on the specific G-SCF polypeptide and the location of the conjugatable attachment groups. For instance, the polypeptide conjugate comprises one or two non-conjugated, conjugatable attachment groups, and at least one, and preferably two or more conjugated attachment groups.

<u>Conjugate of the invention, wherein the non-polypeptide moiety is attached to a lysine or the N-terminal amino acid residue</u>

In one aspect the invention relates to a polypeptide conjugate comprising

i) a polypeptide exhibiting G-CSF activity, comprising an amino acid sequence that differs from the amino acid sequence shown in SEQ ID NO:1 in at least one substitution selected from the group consisting of T1K, P2K, L3K, G4K, P5K, A6K, S7K, S8K, L9K, P10K, Q11K, S12K, F13K, L14K, L15K, E19K, Q20K, V21K, Q25K, G26K, D27K, A29K, A30K, E33K, A37K, T38K, Y39K, L41K, H43K, P44K, E45K, E46K, V48K, L49K, L50K, H52K, S53K, L54K, I56K, P57K, P60K, L61K, S62K, S63K, P65K, S66K, Q67K, A68K, L69K, Q70K, L71K, A72K, G73K, S76K, Q77K, L78K, S80K, F83K, Q86K, G87K, Q90K, E93K, G94K, S96K, P97K, E98K, L99K, G100K, P101K, T102K, D104K, T105K, Q107K, L108K, D109K, A111K, D112K, F113K, T115K, T116K, W118K, Q119K, Q120K, M121K, E122K,

E123K, L124K, M126K, A127K, P128K, A129K, L130K, Q131K, P132K, T133K, Q134K, G135K, A136K, M137K, P138K, A139K, A141K, S142K, A143K, F144K, Q145K, S155K, H156K, Q158K, S159K, L161K, E162K, V163K, S164K, Y165K, V167K, L168K, H170K, L171K, A172K, Q173K and P174K, and

ii) at least one non-polypeptide moiety attached to a lysine residue of the polypeptide.

hG-CSF contains four lysine residues, of which K16 is located in the receptor-binding domain and the others are located in positions 23, 34 and 40, respectively, all relatively close to the receptor-binding domain. In order to avoid conjugation to one or more of these lysine residues (since this may inactivate or severely reduce the activity of the resulting conjugate) it may be desirable to remove at least one lysine residue, e.g., two, three or all of these residues. Accordingly, in another, more preferred aspect the invention relates to a polypeptide conjugate as defined above, wherein at least one of the amino acid residues selected from the group consisting of K16, K23, K34 and K40 has been deleted or substituted with another amino acid residue. Preferably, at least K16 is substituted with another amino acid residue.

Examples of preferred amino acid substitutions include one or more of Q70K, Q90K, T105K, Q120K and T133K, such as two, three or four of these substitutions, for example: Q70K+Q90K, Q70K+T105K, Q70K+120K, Q70K+T133K, Q90K+T105K, Q90K+Q120K, Q90K+T133K, T105K+Q120K, T105K+T133K, Q120K+T133K, Q70K+Q90K+T105K, Q70K+Q90K+Q120K, Q70K+Q90K+T133K, Q70K+T105K+T120K, Q70K+T105K+T133K, Q70K+Q120K+T133K, Q09K+T105K+Q120K, Q90K+T105K+T133K, Q90K+Q120K+T133K, T105K+T120K+T133K, Q90K+T105K+Q120K+T133K, Q70K+T105K+Q120K+T133K, Q70K+Q90K+Q120K+T133K, Q70K+Q90K+T105K+T133K or Q70K+Q90K+T105K+Q120K.

The polypeptide of the conjugate according to the second of the aspects described in the present section (i.e., at least one introduced and one removed lysine) preferably comprises at least one, such as one, two, three or four, of the substitutions selected from the group consisting of K16R, K16Q, K23R, K23Q, K34R, K34Q, K40R and K40Q, more preferably at least one of the substitutions K16R and K23R, whereby conjugation of these residues can be

avoided. Preferably, the polypeptide comprises at least one substitution selected from the group consisting of K16R+K23R, K16R+K34R, K16R+K40R, K23R+K34R, K23R+K40R, K34R+K40R, K16R+K23R+K34R, K16R+K23R+K40R, K23R+K34R+K40R, K16R+K34R+K40R and K16R+K23R+K34R+K40R. These substitutions are likely to give rise to the least structural difference compared to the native or parent polypeptide.

In addition to the substitutions listed above, the polypeptide conjugate of the invention may also include one or more substitutions selected from R22K, R146K, R147K, R166K and R169K.

While the non-polypeptide moiety of the conjugate according to this aspect of the invention may be any molecule which, when using the given conjugation method has lysine as an attachment group such as a carbohydrate moiety, it is preferred that the non-polypeptide moiety is a polymer molecule. The polymer molecule may be any of the molecules mentioned in the section entitled "Conjugation to a polymer molecule", but is preferably selected from the group consisting of linear or branched polyethylene glycol or another polyalkylene oxide. Preferred polymer molecules include, e.g., mPEG-SPA, e.g., from Shearwater Polymers, Inc. or oxycarbonyl-oxy-N-dicarboxyimide PEG (US 5,122,614).

It will be understood that any of the amino acid changes, in particular substitutions, specified in this section can be combined with any of the amino acid changes, prefererably substitutions, specified in the other sections herein disclosing specific amino acid modifications, including introduction and/or removal of glycosylation sites.

<u>Conjugate of the invention, wherein the non-polypeptide moiety is a molecule which has cysteine as an attachment group</u>

In another aspect the invention relates to a conjugate comprising

i) a polypeptide exhibiting G-CSF activity, which comprises an amino acid sequence that differs from the amino acid sequence of hG-CSF shown in SEQ ID NO:1 in at least one substitution selected from the group consisting of T1C, P2C, L3C, G4C, P5C, A6C, S7C, S8C, L9C, P10C, Q11C, S12C, F13C, L14C, L15C, E19C, Q20C, V21C, R22C, Q25C, G26C, D27C, A29C, A30C, E33C, A37C, T38C, Y39C, L41C, H43C, P44C, E45C, E46C, V48C, L49C, L50C, H52C, S53C, L54C, I56C, P57C, P60C, L61C, S62C, S63C, P65C, S66C, Q67C, A68C, L69C, Q70C, L71C, A72C, G73C, S76C, Q77C, L78C, S80C, F83C, Q86C, G87C, Q90C, E93C, G94C, S96C, P97C, E98C, L99C, G100C, P101C, T102C, D104C, T105C, Q107C, L108C, D109C, A111C, D112C, F113C, T115C, T116C, W118C, Q119C, Q120C,

M121C, E122C, E123C, L124C, M126C, A127C, P128C, A129C, L130C, Q131C, P132C, T133C, Q134C, G135C, A136C, M137C, P138C, A139C, A141C, S142C, A143C, F144C, Q145C, R146C, R147C, S155C, H156C, Q158C, S159C, L161C, E162C, V163C, S164C, Y165C, R166C, V167C, L168C, R169C, H170C, L171C, A172C, Q173C and P174C, and

ii) at least one non-polypeptide moiety attached to a cysteine residue of the polypeptide.

The receptor-binding domain of hG-CSF contains a cysteine residue in position 17 which does not take part in a cystine and which may advantageously be removed in order to avoid conjugation of a non-polypeptide moiety to said cysteine. Accordingly, in another, more preferred aspect the invention relates to a conjugate comprising

i) a polypeptide exhibiting G-CSF activity, which comprises an amino acid sequence that differs from the amino acid sequence shown in SEQ ID NO:1 in at least one substitution selected from the group consisting of T1C, P2C, L3C, G4C, P5C, A6C, S7C, S8C, L9C, P10C, Q11C, S12C, F13C, L14C, L15C, E19C, Q20C, V21C, R22C, Q25C, G26C, D27C, A29C, A30C, E33C, A37C, T38C, Y39C, L41C, H43C, P44C, E45C, E46C, V48C, L49C, L50C, H52C, S53C, L54C, I56C, P57C, P60C, L61C, S62C, S63C, P65C, S66C, Q67C, A68C, L69C, Q70C, L71C, A72C, G73C, S76C, Q77C, L78C, S80C, F83C, Q86C, G87C, Q90C, E93C, G94C, S96C, P97C, E98C, L99C, G100C, P101C, T102C, D104C, T105C, Q107C, L108C, D109C, A111C, D112C, F113C, T115C, T116C, W118C, Q119C, Q120C, M121C, E122C, E123C, L124C, M126C, A127C, P128C, A129C, L130C, Q131C, P132C, T133C, Q134C, G135C, A136C, M137C, P138C, A139C, A141C, S142C, A143C, F144C, Q145C, R146C, R147C, S155C, H156C, Q158C, S159C, L161C, E162C, V163C, S164C, Y165C, R166C, V167C, L168C, R169C, H170C, L171C, A172C, Q173C and P174C, in combination with removal of C17, preferably substitution of C17 with any other amino acid residue, e.g., with a serine residue, and

ii) a non-polypeptide moiety which has a cysteine residue as an attachment group.

Preferred substitutions according to this aspect of the invention are substitutions of arginine with cysteine, for example one or more of R146C, R147C, R166C and R169C.

It will be understood that any of the amino acid modifications, in particular substitutions, specified in this section can be combined with any of the amino acid changes, in

particular substitutions, specified in the other sections herein disclosing specific amino acid modifications, including introduction and/or removal of glycosylation sites.

Conjugate of the invention wherein the non-polypeptide moiety binds to an acid group or the C-terminal amino acid residue

In a still further aspect the invention relates to a conjugate comprising

i) a polypeptide exhibiting G-CSF activity, which comprises an amino acid sequence that differs from the amino acid sequence shown in SEQ ID NO:1 in at least one substitution selected from the group consisting of T1D, P2D, L3D, G4D, P5D, A6D, S7D, S8D, L9D, P10D, Q11D, S12D, F13D, L14D, L15D, K16D, Q20D, V21D, R22D, K23D, Q25D, G26D, A29D, A30D, K34D, A37D, T38D, Y39D, K40D, L41D, H43D, P44D, V48D, L49D, L50D, H52D, S53D, L54D, I56D, P57D, P60D, L61D, S62D, S63D, P65D, S66D, Q67D, A68D, L69D, Q70D, L71D, A72D, G73D, S76D, Q77D, L78D, S80D, F83D, Q86D, G87D, Q90D, G94D, S96D, P97D, L99D, G100D, P101D, T102D, T105D, Q107D, L108D, A111D, F113D, T115D, T116D, W118D, Q119D, Q120D, M121D, L124D, M126D, A127D, P128D, A129D, L130D, Q131D, P132D, T133D, Q134D, G135D, A136D, M137D, P138D, A139D, A141D, S142D, A143D, F144D, Q145D, R146D, R147D, S155D, H156D, Q158D, S159D, L161D, V163D, S164D, Y165D, R166D, V167D, L168D, R169D, H170D, L171D, A172D, Q173D and P174D; or at least one substitution selected from the group consisting of T1E, P2E, L3E, G4E, P5E, A6E, S7E, S8E, L9E, P10E, Q11E, S12E, F13E, L14E, L15E, K16E, Q20E, V21E, R22E, K23E, Q25E, G26E, A29E, A30E, K34E, A37E, T38E, Y39E, K40E, L41E, H43E, P44E, V48E, L49E, L50E, H52E, S53E, L54E, I56E, P57E, P60E, L61E, S62E, S63E, P65E, S66E, Q67E, A68E, L69E, Q70E, L71E, A72E, G73E, S76E, Q77E, L78E, S80E, F83E, Q86E, G87E, Q90E, G94E, S96E, P97E, L99E, G100E, P101E, T102E, T105E, Q107E, L108E, A111E, F113E, T115E, T116E, W118E, Q119E, Q120E, M121E, L124E, M126E, A127E, P128E, A129E, L130E, Q131E, P132E, T133E, Q134E, G135E, A136E, M137E, P138E, A139E, A141E, S142E, A143E, F144E, Q145E, R146E, R147E, S155E, H156E, Q158E, S159E, L161E, V163E, S164E, Y165E, R166E, V167E, L168E, R169E, H170E, L171E, A172E, Q173E and P174E; and

ii) a non-polypeptide moiety having an aspartic acid or a glutamic acid residue as an attachment group.

Examples of preferred substitutions according to this aspect of the invention include Q67D/E, Q70D/E, Q77D/E, Q86D/E, Q90D/E, Q120D/E, Q131D/E, Q134D/E, Q145D/E and Q173D/E.

In addition to the above listed substitutions, the polypeptide of the conjugate according to any of the above aspects may comprise removal, preferably by substitution, of at least one of the amino acid residues selected from the group consisting of D27, D104, D109, D112, E19, E33, E45, E46, E93, E98, E122, E123, and E163. The substitution may be for any other amino acid residue, in particular for an asparagine or a glutamine residue, whereby conjugation of these residues can be avoided. In particular, the polypeptide may comprise at least one of the following substitutions: D27N, D104N, D109N, D112N, E19Q, E33Q, E45Q, E46Q, E93Q, E98Q, E122Q, E123Q and E163Q. Preferably, the amino acid substitution in one or more of the above positions may in addition be combined with at least one of the following substitutions: D109N, D112N, E19Q, E122Q and E123Q. Substitution with any of these amino acid residues is likely to give rise to the least structural difference.

While the non-polypeptide moiety of the conjugate according to this aspect of the invention, which has an acid group as an attachment group, can be any non-polypeptide moiety with such property, it is presently preferred that the non-polypeptide moiety is a polymer molecule or an organic derivatizing agent, in particular a polymer molecule, and the conjugate is prepared e.g., as described by Sakane and Pardridge, Pharmaceutical Research, Vol. 14, No. 8, 1997, pp 1085-1091.

It will be understood that any of the amino acid changes, in particular substitutions, specified in this section can be combined with any of the amino acid changes, in particular substitutions specified in the other sections herein disclosing specific amino acid changes, including introduction and/or removal of glycosylation sites.

Other conjugates of the invention

In addition to the non-polypeptide moieties specified above e.g., in the sections entitled "Conjugate of the invention...." the conjugate of the invention may contain further carbohydrate moieties as a consequence of the polypeptide being expressed in a glycosylating host cell and glycosylation achieved at the glycosylation sites or introduced glycosylation site(s).

Conjugate of the invention wherein the non-polypeptide moiety is a carbohydrate moiety

In a further aspect the invention relates to a conjugate comprising a glycosylated polypeptide exhibiting G-CSF activity, which comprises an amino acid sequence that differs from that shown in SEQ ID NO:1 in that at least one non-naturally occurring glycosylation site has been introduced into the amino acid sequence by way of at least one substitution selected from the group consisting of L3N+P5S/T, P5N, A6N, S8N+P10S/T, P10N, Q11N+F13S/T, S12N+L14S/T, F13N+L15S/T, L14N+K16S/T, K16N+L18S/T, E19N+V21S/T, Q20N+R22S/T, V21N+K23S/T, R22N+I24S/T, K23N+Q25S/T, Q25N+D27S/T, G26N+G28S/T, D27N+A29S/T, A29N+L31S/T, A30N+Q32S/T, E33N+L35S/T, A37N+Y39S/T, T38N+K40S/T, Y39N+L41S/T, P44N+E46S/T, E45N+L47S/T, E46N+V48S/T, V48N+L50S/T, L49N+G51S/T, L50N+H52S/T, H52N+L54S/T, S53N+G55S/T, P60N, L61N, S63N+P65S/T, P65N+Q67S/T, S66N+A68S/T, Q67N+L69S/T, A68N+Q70S/T, L69N+ L71S/T, Q70N+A72S/T, L71N+G73S/T, G73N+L75S/T, S76N+L78S/T, Q77N+H79S/T, L78N, S80N+L82S/T, F83N+Y85S/T, Q86N+L88S/T, G87N+L89S/T, Q90N+L92S/T, E93N+I95S/T, P97N+L99S/T, L99N+P101S/T, P101N+L103S/T, T102N+D104S/T, D104N+L106S/T, T105N+Q107S/T, Q107N+D109S/T, L108N+V110S/T, D109N+A111S/T, A111N+F113S/T, D112N+A114S/T, F113N, T115N+I117S/T, T116N+W118S/T, W118N+Q120S/T, Q119N+M121S/T, Q120N+E122S/T, M121N+E123S/T, E122N+L124S/T, E123N+G125S/T, L124N+M126S/T, M126N+P128S/T, P128N+L130S/T, L130N+P132S/T, P132N+Q134S/T, T133N+G135S/T, Q134N+A136S/T, A136N+P138S/T, P138N+F140S/T, A139N+A141S/T, A141N+A143S/T, S142N+F144S/T, A143N+Q145S/T, F144N+R146S/T, Q145N+R147S/T, R146N+A148S/T, R147N+G149S/T, S155N+L157S/T, H156N+Q158S/T, S159N+L161S/T, L161N+V163S/T, E162N, V163N+Y165S/T, S164N+R166S/T, Y165N+V167S/T, R166N+L168S/T, V167N+R169S/T, L168N+H170S/T, R169N+L171S/T and H170N+A172S/T, wherein S/T indicates an S or a T residue, preferably a T residue.

It will be understood that in order to prepare a conjugate according to this aspect the polypeptide must be expressed in a glycosylating host cell capable of attaching oligosaccharide moieties at the glycosylation site(s) or alternatively subjected to *in vitro* glycosylation. Examples of glycosylating host cells are given in the section further below entitled "Coupling to an oligosaccharide moiety".

Alternatively, the conjugate according to this aspect comprises a polypeptide exhibiting G-CSF activity, which comprises an amino acid sequence that differs from that shown in SEQ ID NO:1 in at least one substitution selected from the group consisting of P5N, A6N, P10N, P60N, L61N, L78N, F113N and E162N, in particular from the group consisting of P5N, A6N, P10N, P60N, L61N, F113N and E162N, such as from the group consisting of P60N, L61N, F113N and E162N.

Alternatively, the conjugate according to this aspect comprises a polypeptide exhibiting G-CSF activity, which comprises an amino acid sequence that differs from that shown in SEQ ID NO:1 in at least one substitution selected from the group consisting of D27N+A29S, D27N+A29T, D104N+L106S, D104N+L106T, D109N+A111S, D109N+A111T, D112N+A114S and D112N+A114T, more preferably from the group consisting of D27N+A29S, D27N+A29T, D104N+L106S, D104N+L106T, D112N+A114S and D112N+A114T, such as from the group consisting of D27N+A29S, D27N+A29T, D104N+L106S and D104N+L106T.

In addition to a carbohydrate molecule, the conjugate according to the aspect of the invention described in the present section may contain additional non-polypeptide moieties, in particular a polymer molecule, as described in the present application, conjugated to one or more attachment groups present in the polypeptide part of the conjugate.

It will be understood that any of the amino acid changes, in particular substitutions, specified in this section can be combined with any of the amino acid changes, in particular substitutions, specified in the other sections herein disclosing specific amino acid changes.

Circularly permuted variants

In a further embodiment, the polypeptide part of the polypeptide conjugate of the invention may be in the form of a circularly permuted variant of a polypeptide sequence otherwise disclosed herein. In such a circularly permuted polypeptide, the original N-terminus and C-terminus are joined together either directly by a peptide bond or indirectly via a peptide linker, while new N- and C-termini are formed between two adjacent amino acid residues that originally were joined by a peptide bond. Since the original N- and C-termini will normally be located at some distance from each other, they will typically be linked by means of a peptide linker having a suitable length and composition so that the structure and activity of the conjugate is not adversely affected. It will be clear that the new N-terminus and C-terminus should not be formed between an amino acid residue pair where this would interfere with the activity of the

polypeptide. Circularly permuted G-CSF receptor agonists are disclosed in US 6,100,070, to which reference is made for further information on selecting peptide linkers and the location of the new N-terminus and C-terminus as well as methods for producing them such variants.

White blood cell and neutrophil formation of conjugates of the invention

In a further embodiment, the polypeptide conjugate of the invention may be characterized as being a conjugate exhibiting G-CSF activity and comprising a polypeptide with an amino acid sequence that differs in at least one amino acid residue from the amino acid sequence shown in SEQ ID NO:1 and having at least one non-polypeptide moiety attached to an attachment group of the polypeptide, the polypeptide conjugate further fulfilling at least one of the following criteria (A)-(D):

(A) after one subcutaneous administration of 100 microgram per kg body weight to rats (based on the weight of the polypeptide part of the conjugate) it:

i) increases formation of white blood cells with at least about the same rate and to at least about the same level (measured as number of cells per liter of blood) as administration of 100 microgram of non-conjugated hG-CSF per kg body weight for a period of 12 hours after administration, and

ii) increases the level of white blood cells (measured as number of cells per liter blood) above the level of white blood cells prior to administration for a period of at least about 96 hours, preferably for at least about 120 hours;

(B) after one subcutaneous administration of 25 microgram per kg body weight to rats (based on the weight of the polypeptide part of the conjugate) it:

i) increases formation of white blood cells with at least about the same rate and to at least about the same level (measured as number of cells per liter of blood) as administration of 100 microgram of non-conjugated hG-CSF per kg body weight for a period of 12 hours after administration, and

ii) increases the level of white blood cells (measured as number of cells per liter blood) above the level of white blood cells prior to administration for a period of at least about 72 hours, preferably at least about 96 hours, more preferably at least about 120 hours;

(C) after one subcutaneous administration of 100 microgram per kg body weight to rats (based on the weight of the polypeptide part of the conjugate) it:

i) increases formation of neutrophils with at least about the same rate and to at least about the same level (measured as number of cells per liter of blood) as administration of 100 microgram of non-conjugated hG-CSF per kg body weight for a period of 12 hours after administration, and

ii) increases the level of neutrophils (measured as number of cells per liter blood) above the level of neutrophils prior to administration for a period of at least about 96 hours, preferably at least about 120 hours;

(D) after one subcutaneous administration of 25 microgram per kg body weight to rats (based on the weight of the polypeptide part of the conjugate) it:

i) increases formation of neutrophils with at least about the same rate and to at least about the same level (measured as number of cells per liter of blood) as administration of 100 microgram of non-conjugated hG-CSF per kg body weight for a period of 12 hours after administration, and

ii) increases the level of neutrophils (measured as number of cells per liter blood) above the level of neutrophils prior to administration for a period of at least about 72 hours, preferably at least about 96 hours, more preferably at least about 120 hours.

Non-polypeptide moiety of the conjugate of the invention

As indicated further above the non-polypeptide moiety of the conjugate of the invention is preferably selected from the group consisting of a polymer molecule, a lipophilic compound, a carbohydrate moiety (e.g., by way of *in vivo* glycosylation) and an organic derivatizing agent. All of these agents may confer desirable properties to the polypeptide part of the conjugate, in particular increased functional *in vivo* half-life and/or increased serum half-life. The polypeptide part of the conjugate is normally conjugated to only one type of non-polypeptide moiety, but may also be conjugated to two or more different types of non-polypeptide moieties, e.g., to a polymer molecule and an oligosaccharide moiety, to a lipophilic group and an oligosaccharide moiety, to an organic derivatizing agent and an oligosaccharide moiety, to a lipophilic group and a polymer molecule, etc. The conjugation to two or more different non-polypeptide moieties may be done simultaneously or sequentially.

Methods for preparing a conjugate of the invention

In the following sections "Conjugation to a lipophilic compound", "Conjugation to a polymer molecule", "Conjugation to an oligosaccharide moiety" and "Conjugation to an

organic derivatizing agent" conjugation to specific types of non-polypeptide moieties is described. In general, a polypeptide conjugate according to the invention may be produced by culturing an appropriate host cell under conditions conducive for expression of the polypeptide, and recovering the polypeptide, wherein a) the polypeptide comprises at least one N- or O-glycosylation site and the host cell is a eukaryotic host cell capable of *in vivo* glycosylation, and/or b) the polypeptide is subjected to conjugation to a non-polypeptide moiety *in vitro*.

Conjugation to a lipophilic compound

The polypeptide and the lipophilic compound may be conjugated to each other, either directly or by use of a linker. The lipophilic compound may be a natural compound such as a saturated or unsaturated fatty acid, a fatty acid diketone, a terpene, a prostaglandin, a vitamin, a carotenoid or steroid, or a synthetic compound such as a carbon acid, an alcohol, an amine and sulphonic acid with one or more alkyl, aryl, alkenyl or other multiple unsaturated compounds. The conjugation between the polypeptide and the lipophilic compound, optionally through a linker, may be done according to methods known in the art, e.g., as described by Bodanszky in Peptide Synthesis, John Wiley, New York, 1976 and in WO 96/12505.

Conjugation to a polymer molecule

The polymer molecule to be coupled to the polypeptide may be any suitable polymer molecule, such as a natural or synthetic homo-polymer or heteropolymer, typically with a molecular weight in the range of about 300-100,000 Da, such as about 500-20,000 Da, more preferably in the range of about 1000-15,000 Da, even more preferably in the range of about 2000-12,000 Da, such as about 3000-10,000. When used about polymer molecules herein, the word "about" indicates an approximate average molecular weight and reflects the fact that there will normally be a certain molecular weight distribution in a given polymer preparation.

Examples of homo-polymers include a polyol (i.e., poly-OH), a polyamine (i.e., poly-NH_2) and a polycarboxylic acid (i.e., poly-COOH). A hetero-polymer is a polymer which comprises different coupling groups, such as a hydroxyl group and an amine group.

Examples of suitable polymer molecules include polymer molecules selected from the group consisting of polyalkylene oxide (PAO), including polyalkylene glycol (PAG), such as linear or branched polyethylene glycol (PEG) and polypropylene glycol (PPG), poly-vinyl alcohol (PVA), poly-carboxylate, poly-(vinylpyrolidone), polyethylene-co-maleic acid anhydride, polystyrene-co-maleic acid anhydride, dextran, including carboxymethyl-dextran, or

any other biopolymer suitable for reducing immunogenicity and/or increasing functional *in vivo* half-life and/or serum half-life. Another example of a polymer molecule is human albumin or another abundant plasma protein. Generally, polyalkylene glycol-derived polymers are biocompatible, non-toxic, non-antigenic, non-immunogenic, have various water solubility properties, and are easily excreted from living organisms.

PEG is the preferred polymer molecule, since it has only few reactive groups capable of cross-linking compared to polysaccharides such as dextran. In particular, monofunctional PEG, e.g., methoxypolyethylene glycol (mPEG), is of interest since its coupling chemistry is relatively simple (only one reactive group is available for conjugating with attachment groups on the polypeptide). Consequently, the risk of cross-linking is eliminated, the resulting polypeptide conjugates are more homogeneous and the reaction of the polymer molecules with the polypeptide is easier to control.

To effect covalent attachment of the polymer molecule(s) to the polypeptide, the hydroxyl end groups of the polymer molecule are provided in activated form, i.e., with reactive functional groups. Suitable activated polymer molecules are commercially available, e.g., from Shearwater Polymers, Inc., Huntsville, AL, USA, or from PolyMASC Pharmaceuticals plc, UK. Alternatively, the polymer molecules can be activated by conventional methods known in the art, e.g., as disclosed in WO 90/13540. Specific examples of activated linear or branched polymer molecules for use in the present invention are described in the Shearwater Polymers, Inc. 1997 and 2000 Catalogs (Functionalized Biocompatible Polymers for Research and pharmaceuticals, Polyethylene Glycol and Derivatives, incorporated herein by reference). Specific examples of activated PEG polymers include the following linear PEGs: NHS-PEG (e.g., SPA-PEG, SSPA-PEG, SBA-PEG, SS-PEG, SSA-PEG, SC-PEG, SG-PEG, and SCM-PEG), and NOR-PEG), BTC-PEG, EPOX-PEG, NCO-PEG, NPC-PEG, CDI-PEG, ALD-PEG, TRES-PEG, VS-PEG, IODO-PEG, and MAL-PEG, and branched PEGs such as PEG2-NHS and those disclosed in US 5,932,462 and US 5,643,575, both of which are incorporated herein by reference. Furthermore, the following publications, incorporated herein by reference, disclose useful polymer molecules and/or PEGylation chemistries: US 5,824,778, US 5,476,653, WO 97/32607, EP 229,108, EP 402,378, US 4,902,502, US 5,281,698, US 5,122,614, US 5,219,564, WO 92/16555, WO 94/04193, WO 94/14758, WO 94/17039, WO 94/18247, WO 94/28024, WO 95/00162, WO 95/11924, WO95/13090, WO 95/33490, WO 96/00080, WO 97/18832, WO 98/41562, WO

98/48837, WO 99/32134, WO 99/32139, WO 99/32140, WO 96/40791, WO 98/32466, WO 95/06058, EP 439 508, WO 97/03106, WO 96/21469, WO 95/13312, EP 921 131, US 5,736,625, WO 98/05363, EP 809 996, US 5,629,384, WO 96/41813, WO 96/07670, US 5,473,034, US 5,516,673, EP 605 963, US 5,382,657, EP 510 356, EP 400 472, EP 183 503 and EP 154 316.

The conjugation of the polypeptide and the activated polymer molecules is conducted by use of any conventional method, e.g., as described in the following references (which also describe suitable methods for activation of polymer molecules): R.F. Taylor, (1991), "Protein immobilisation. Fundamental and applications", Marcel Dekker, N.Y.; S.S. Wong, (1992), "Chemistry of Protein Conjugation and Crosslinking", CRC Press, Boca Raton; G.T. Hermanson et al., (1993), "Immobilized Affinity Ligand Techniques", Academic Press, N.Y.). The skilled person will be aware that the activation method and/or conjugation chemistry to be used depends on the attachment group(s) of the polypeptide (examples of which are given further above), as well as the functional groups of the polymer (e.g., being amine, hydroxyl, carboxyl, aldehyde, sulfydryl, succinimidyl, maleimide, vinysulfone or haloacetate). The PEGylation may be directed towards conjugation to all available attachment groups on the polypeptide (i.e., such attachment groups that are exposed at the surface of the polypeptide) or may be directed towards one or more specific attachment groups, e.g., the N-terminal amino group (US 5,985,265). Furthermore, the conjugation may be achieved in one step or in a stepwise manner (e.g., as described in WO 99/55377).

It will be understood that the PEGylation is designed so as to produce the optimal molecule with respect to the number of PEG molecules attached, the size and form of such molecules (e.g., whether they are linear or branched), and where in the polypeptide such molecules are attached. The molecular weight of the polymer to be used will be chosen taking into consideration the desired effect to be achieved. For instance, if the primary purpose of the conjugation is to achieve a conjugate having a high molecular weight and larger size (e.g., to reduce renal clearance), one may choose to conjugate either one or a few high molecular weight polymer molecules or a number of polymer molecules with a smaller molecular weight to obtain the desired effect. Preferably, however, several polymer molecules with a smaller molecular weight will be used. When a high degree of epitope shielding is desirable, this may be obtained by use of a sufficiently high number of low molecular weight polymer molecules (e.g., with a molecular weight of about 5,000 Da) to effectively shield all or most epitopes of the polypeptide.

For instance, 2-8, such as 3-6 such polymers may be used. As the examples below illustrate, it may be advantageous to have a larger number of polymer molecules with a lower molecular weight (e.g., 4-6 with a MW of 5000) compared to a smaller number of polymer molecules with a higher molecular weight (e.g., 1-3 with a MW of 12,000-20,000) in terms of improving the functional *in vivo* half-life of the polypeptide conjugate, even where the total molecular weight of the attached polymer molecules in the two cases is the same. It is believed that the presence of a larger number of smaller polymer molecules provides the polypeptide with a larger diameter or apparent size than e.g., a single yet larger polymer molecule, at least when the polymer molecules are relatively uniformly distributed on the polypeptide surface.

While conjugation of only a single polymer molecule to a single attachment group on the protein is not preferred, in the event that only one polymer molecule is attached, it will generally be advantageous that the polymer molecule, which may be linear or branched, has a relatively high molecular weight, e.g., about 20 kDa.

Normally, the polymer conjugation is performed under conditions aiming at reacting as many of the available polymer attachment groups as possible with polymer molecules. This is achieved by means of a suitable molar excess of the polymer in relation to the polypeptide. Typical molar ratios of activated polymer molecules to polypeptide are up to about 1000-1, such as up to about 200-1 or up to about 100-1. In some cases, the ratio may be somewhat lower, however, such as up to about 50-1, 10-1 or 5-1.

It is also contemplated according to the invention to couple the polymer molecules to the polypeptide through a linker. Suitable linkers are well known to the skilled person. A preferred example is cyanuric chloride (Abuchowski et al., (1977), J. Biol. Chem., 252, 3578-3581; US 4,179,337; Shafer et al., (1986), J. Polym. Sci. Polym. Chem. Ed., 24, 375-378.

Subsequent to the conjugation residual activated polymer molecules are blocked according to methods known in the art, e.g., by addition of primary amine to the reaction mixture, and the resulting inactivated polymer molecules are removed by a suitable method (see Materials and Methods).

In a preferred embodiment, the polypeptide conjugate of the invention comprises a PEG molecule attached to some, most or preferably substantially all of the lysine residues in the polypeptide available for PEGylation, in particular a linear or branched PEG molecule, e.g., with

a molecular weight of about 1-15 kDa, typically about 2-12 kDa, such as about 3-10 kDa, e.g., about 5 or 6 kDa.

It will be understood that depending on the circumstances, e.g., the amino acid sequence of the polypeptide, the nature of the activated PEG compound being used and the specific PEGylation conditions, including the molar ratio of PEG to polypeptide, varying degrees of PEGylation may be obtained, with a higher degree of PEGylation generally being obtained with a higher ratio of PEG to polypeptide. The PEGylated polypeptides resulting from any given PEGylation process will, however, normally comprise a stochastic distribution of polypeptide conjugates having slightly different degrees of PEGylation.

In yet another embodiment, the polypeptide conjugate of the invention may comprise a PEG molecule attached to the lysine residues in the polypeptide available for PEGylation, and in addition to the N-terminal amino acid residue of the polypeptide.

Coupling to an oligosaccharide moiety

The conjugation to an oligosaccharide moiety may take place *in vivo* or *in vitro*. In order to achieve *in vivo* glycosylation of a G-CSF molecule comprising one or more glycosylation sites the nucleotide sequence encoding the polypeptide must be inserted in a glycosylating, eucaryotic expression host. The expression host cell may be selected from fungal (filamentous fungal or yeast), insect or animal cells or from transgenic plant cells. In one embodiment the host cell is a mammalian cell, such as a CHO cell, BHK or HEK, e.g., HEK 293, cell, or an insect cell, such as an SF9 cell, or a yeast cell, e.g., *S. cerevisiae* or *Pichia pastoris*, or any of the host cells mentioned hereinafter. Covalent *in vitro* coupling of glycosides (such as dextran) to amino acid residues of the polypeptide may also be used, e.g., as described in WO 87/05330 and in Aplin et al., CRC Crit Rev. Biochem., pp. 259-306, 1981.

The *in vitro* coupling of oligosaccharide moieties or PEG to protein- and peptide-bound Gln-residues can be carried out by transglutaminases (TG'ases). Transglutaminases catalyse the transfer of donor amine-groups to protein- and peptide-bound Gln-residues in a so-called cross-linking reaction. The donor-amine groups can be protein- or peptide-bound e.g., as the ε-amino-group in Lys-residues or can be part of a small or large organic molecule. An example of a small organic molecule functioning as an amino-donor in TG'ase-catalysed cross-linking is putrescine (1,4-diaminobutane). An example of a larger organic molecule functioning

as an amino-donor in TG'ase-catalysed cross-linking is an amine-containing PEG (Sato et al., Biochemistry 35, 13072-13080).

Tg'ases are in general highly specific enzymes, and not every Gln-residue exposed on the surface of a protein is accessible to TG'ase-catalysed cross-linking to amino-containing substances. On the contrary, only a few Gln-residues function naturally as TG'ase substrates, but the exact parameters governing which Gln-residues are good TG'ase substrates remain unknown. Thus, in order to render a protein susceptible to TG'ase-catalysed cross-linking reactions it is often a prerequisite to add at convenient positions stretches of amino acid sequence known to function very well as TG'ase substrates. Several amino acid sequences are known to be or to contain excellent natural TG'ase substrates e.g., substance P, elafin, fibrinogen, fibronectin, α_2-plasmin inhibitor, α-caseins, and β-caseins.

Coupling to an organic derivatizing agent

Covalent modification of the polypeptide exhibiting G-CSF activity may be performed by reacting one or more attachment groups of the polypeptide with an organic derivatizing agent. Suitable derivatizing agents and methods are well known in the art. For example, cysteinyl residues most commonly are reacted with α-haloacetates (and corresponding amines), such as chloroacetic acid or chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteinyl residues also are derivatized by reaction with bromotrifluoroacetone, α-bromo-β-(4-imidozoyl)propionic acid, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole. Histidyl residues are derivatized by reaction with diethylpyrocarbonateat pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-bromophenacyl bromide is also useful. The reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0. Lysinyl and amino terminal residues are reacted with succinic or other carboxylic acid anhydrides. Derivatization with these agents has the effect of reversing the charge of the lysinyl residues. Other suitable reagents for derivatizing α-amino-containing residues include imidoesters such as methyl picolinimidate, pyridoxal phosphate, pyridoxal, chloroborohydride, trinitrobenzenesulfonic acid, O-methylisourea, 2,4-pentanedione and transaminase-catalyzed reaction with glyoxylate. Arginyl residues are modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and

ninhydrin. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pKa of the guanidine functional group.

Furthermore, these reagents may react with the groups of lysine as well as the arginine guanidino group. Carboxyl side groups (aspartyl or glutamyl) are selectively modified by reaction with carbodiimides (R-N=C=N-R'), where R and R' are different alkyl groups, such as 1-cyclohexyl-3-(2-morpholinyl-4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore, aspartyl and glutamyl residues are converted to asparaginyl and glutaminyl residues by reaction with ammonium ions.

Blocking of the functional site

It has been reported that excessive polymer conjugation can lead to a loss of activity of the polypeptide to which the polymer is conjugated. This problem can be eliminated by e.g., removal of attachment groups located at the functional site or by blocking the functional site prior to conjugation so that the functional site is blocked during conjugation. The latter strategy constitutes a further embodiment of the invention (the first strategy being exemplified further above, e.g., by removal of lysine residues which may be located close to the functional site). More specifically, according to the second strategy the conjugation between the polypeptide and the non-polypeptide moiety is conducted under conditions where the functional site of the polypeptide is blocked by a helper molecule capable of binding to the functional site of the polypeptide.

Preferably, the helper molecule is one which specifically recognizes a functional site of the polypeptide, such as a receptor, in particular the G-CSF receptor or a part of the G-CSF receptor.

Alternatively, the helper molecule may be an antibody, in particular a monoclonal antibody recognizing the polypeptide exhibiting G-CSF activity. In particular, the helper molecule may be a neutralizing monoclonal antibody.

The polypeptide is allowed to interact with the helper molecule before effecting conjugation. This ensures that the functional site of the polypeptide is shielded or protected and consequently unavailable for derivatization by the non-polypeptide moiety such as a polymer. Following its elution from the helper molecule, the conjugate between the non-polypeptide moiety and the polypeptide can be recovered with at least a partially preserved functional site.

The subsequent conjugation of the polypeptide having a blocked functional site to a polymer, a lipophilic compound, an oligosaccharide moiety, an organic derivatizing agent or any other compound is conducted in the normal way, e.g., as described in the sections above entitled "Conjugation to ....".

Irrespective of the nature of the helper molecule to be used to shield the functional site of the polypeptide from conjugation, it is desirable that the helper molecule is free of or comprises only a few attachment groups for the non-polypeptide moiety of choice in part(s) of the molecule where the conjugation to such groups would hamper desorption of the conjugated polypeptide from the helper molecule. Hereby, selective conjugation to attachment groups present in non-shielded parts of the polypeptide can be obtained and it is possible to reuse the helper molecule for repeated cycles of conjugation. For instance, if the non-polypeptide moiety is a polymer molecule such as PEG, which has the epsilon amino group of a lysine or N-terminal amino acid residue as an attachment group, it is desirable that the helper molecule is substantially free of conjugatable epsilon amino groups, preferably free of any epsilon amino groups. Accordingly, in a preferred embodiment the helper molecule is a protein or peptide capable of binding to the functional site of the polypeptide, which protein or peptide is free of any conjugatable attachment groups for the non-polypeptide moiety of choice.

Of particular interest in connection with the embodiment of the present invention wherein the polypeptide conjugates are prepared from a diversified population of nucleotide sequences encoding a polypeptide of interest, the blocking of the functional group is effected in microtiter plates prior to conjugation, for instance by plating the expressed polypeptide variant in a microtiter plate containing an immobilized blocking group such as a receptor, an antibody or the like.

In a further embodiment the helper molecule is first covalently linked to a solid phase such as column packing materials, for instance Sephadex or agarose beads, or a surface, e.g., a reaction vessel. Subsequently, the polypeptide is loaded onto the column material carrying the helper molecule and conjugation carried out according to methods known in the art, e.g., as described in the sections above entitled "Conjugation to ....". This procedure allows the polypeptide conjugate to be separated from the helper molecule by elution. The polypeptide conjugate is eluted by conventional techniques under physico-chemical conditions that do not lead to a substantive degradation of the polypeptide conjugate. The fluid phase containing the

polypeptide conjugate is separated from the solid phase to which the helper molecule remains covalently linked. The separation can be achieved in other ways: For instance, the helper molecule may be derivatised with a second molecule (e.g., biotin) that can be recognized by a specific binder (e.g., streptavidin). The specific binder may be linked to a solid phase, thereby allowing the separation of the polypeptide conjugate from the helper molecule-second molecule complex through passage over a second helper-solid phase column which will retain, upon subsequent elution, the helper molecule-second molecule complex, but not the polypeptide conjugate. The polypeptide conjugate may be released from the helper molecule in any appropriate fashion. Deprotection may be achieved by providing conditions in which the helper molecule dissociates from the functional site of the G-CSF to which it is bound. For instance, a complex between an antibody to which a polymer is conjugated and an anti-idiotypic antibody can be dissociated by adjusting the pH to an acid or alkaline pH.

Conjugation of a tagged polypeptide

In an alternative embodiment the polypeptide is expressed as a fusion protein with a tag, i.e., an amino acid sequence or peptide stretch made up of typically 1-30, such as 1-20 amino acid residues. Besides allowing for fast and easy purification, the tag is a convenient tool for achieving conjugation between the tagged polypeptide and the non-polypeptide moiety. In particular, the tag may be used for achieving conjugation in microtiter plates or other carriers, such as paramagnetic beads, to which the tagged polypeptide can be immobilised via the tag. The conjugation to the tagged polypeptide in e.g., microtiter plates has the advantage that the tagged polypeptide can be immobilised in the microtiter plates directly from the culture broth (in principle without any purification) and subjected to conjugation. Thereby, the total number of process steps (from expression to conjugation) can be reduced. Furthermore, the tag may function as a spacer molecule, ensuring an improved accessibility to the immobilised polypeptide to be conjugated. The conjugation using a tagged polypeptide may be to any of the non-polypeptide moieties disclosed herein, e.g., to a polymer molecule such as PEG.

The identity of the specific tag to be used is not critical as long as the tag is capable of being expressed with the polypeptide and is capable of being immobilised on a suitable surface or carrier material. A number of suitable tags are commercially available, e.g., from Unizyme Laboratories, Denmark. For instance, the tag may consist of any of the following sequences:

His-His-His-His-His-His

Met-Lys-His-His-His-His-His-His

Met-Lys-His-His-Ala-His-His-Gln-His-His

Met-Lys-His-Gln-His-Gln-His-Gln-His-Gln-His-Gln-His-Gln

Met-Lys-His-Gln-His-Gln-His-Gln-His-Gln-His-Gln-His-Gln-Gln

or any of the following:

EQKLI SEEDL (a C-terminal tag described in Mol. Cell. Biol. 5:3610-16, 1985)

DYKDDDDK (a C- or N-terminal tag)

YPYDVPDYA

Antibodies against the above tags are commercially available, e.g., from ADI, Aves Lab and Research Diagnostics.

A convenient method for using a tagged polypeptide for PEGylation is given in the Materials and Methods section below. The subsequent cleavage of the tag from the polypeptide may be achieved by use of commercially available enzymes.

Methods for preparing a polypeptide of the invention or the polypeptide part of the conjugate of the invention

The polypeptide of the present invention or the polypeptide part of a conjugate of the invention, optionally in glycosylated form, may be produced by any suitable method known in the art. Such methods include constructing a nucleotide sequence encoding the polypeptide and expressing the sequence in a suitable transformed or transfected host. However, polypeptides of the invention may be produced, albeit less efficiently, by chemical synthesis or a combination of chemical synthesis or a combination of chemical synthesis and recombinant DNA technology.

A nucleotide sequence encoding a polypeptide or the polypeptide part of a conjugate of the invention may be constructed by isolating or synthesizing a nucleotide sequence encoding the parent G-CSF, such as hG-CSF with the amino acid sequence shown in SEQ ID NO:1, and then changing the nucleotide sequence so as to effect introduction (i.e., insertion or substitution) or deletion (i.e., removal or substitution) of the relevant amino acid residue(s).

The nucleotide sequence is conveniently modified by site-directed mutagenesis in accordance with conventional methods. Alternatively, the nucleotide sequence is prepared by chemical synthesis, e.g., by using an oligonucleotide synthesizer, wherein oligonucleotides are designed based on the amino acid sequence of the desired polypeptide, and preferably selecting those codons that are favored in the host cell in which the recombinant polypeptide will be

produced. For example, several small oligonucleotides coding for portions of the desired polypeptide may be synthesized and assembled by PCR, ligation or ligation chain reaction (LCR) (Barany, PNAS 88:189-193, 1991). The individual oligonucleotides typically contain 5' or 3' overhangs for complementary assembly.

Alternative nucleotide sequence modification methods are available for producing polypeptide variants for high throughput screening, for instance methods which involve homologous cross-over such as disclosed in US 5,093,257, and methods which involve gene shuffling, i.e., recombination between two or more homologous nucleotide sequences resulting in new nucleotide sequences having a number of nucleotide alterations when compared to the starting nucleotide sequences. Gene shuffling (also known as DNA shuffling) typically involves one or more cycles of random fragmentation and reassembly of the nucleotide sequences, followed by screening to select nucleotide sequences encoding polypeptides with desired properties. In order for homology-based nucleic acid shuffling to take place, the relevant parts of the nucleotide sequences are preferably at least 50% identical, such as at least 60% identical, more preferably at least 70% identical, such as at least 80% identical. The recombination can be performed *in vitro* or *in vivo*.

Examples of suitable *in vitro* gene shuffling methods are disclosed by Stemmer et al. (1994), Proc. Natl. Acad. Sci. USA; vol. 91, pp. 10747-10751; Stemmer (1994), Nature, vol. 370, pp. 389-391; Smith (1994), Nature vol. 370, pp. 324-325; Zhao et al., Nat. Biotechnol. 1998, Mar; 16(3): 258-61; Zhao H. and Arnold, FB, Nucleic Acids Research, 1997, Vol. 25. No. 6 pp. 1307-1308; Shao et al., Nucleic Acids Research 1998, Jan 15; 26(2): pp. 681-83; and WO 95/17413. An example of a suitable *in vivo* shuffling method is disclosed in WO 97/07205. Other techniques for mutagenesis of nucleic acid sequences by *in vitro* or *in vivo* recombination are disclosed e.g., in WO 97/20078 and US 5,837,458. Examples of specific shuffling techniques include "family shuffling", "synthetic shuffling" and "*in silico* shuffling". Family shuffling involves subjecting a family of homologous genes from different species to one or more cycles of shuffling and subsequent screening or selection. Family shuffling techniques are disclosed e.g., by Crameri et al. (1998), Nature, vol. 391, pp. 288-291; Christians et al. (1999), Nature Biotechnology, vol. 17, pp. 259-264; Chang et al. (1999), Nature Biotechnology, vol. 17, pp. 793-797; and Ness et al. (1999), Nature Biotechnology, vol. 17, 893-896. Synthetic shuffling involves providing libraries of overlapping synthetic oligonucleotides based e.g., on a sequence

alignment of homologous genes of interest. The synthetically generated oligonucleotides are recombined, and the resulting recombinant nucleic acid sequences are screened and if desired used for further shuffling cycles. Synthetic shuffling techniques are disclosed in WO 00/42561. *In silico* shuffling refers to a DNA shuffling procedure which is performed or modelled using a computer system, thereby partly or entirely avoiding the need for physically manipulating nucleic acids. Techniques for *in silico* shuffling are disclosed in WO 00/42560.

Once assembled (by synthesis, site-directed mutagenesis or another method), the nucleotide sequence encoding the polypeptide is inserted into a recombinant vector and operably linked to control sequences necessary for expression of the G-CSF in the desired transformed host cell.

It should of course be understood that not all vectors and expression control sequences function equally well to express the nucleotide sequence encoding a polypeptide described herein. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation. For example, in selecting a vector, the host must be considered because the vector must replicate in it or be able to integrate into the chromosome. The vector's copy number, the ability to control that copy number, and the expression of any other proteins encoded by the vector, such as antibiotic markers, should also be considered. In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleotide sequence encoding the polypeptide, particularly as regards potential secondary structures. Hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleotide sequence, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification of the products coded for by the nucleotide sequence.

The recombinant vector may be an autonomously replicating vector, i.e., a vector, which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g., a plasmid. Alternatively, the vector is one which, when introduced into a host cell, is integrated into the host cell genome and replicated together with the chromosome(s) into which it has been integrated.

The vector is preferably an expression vector in which the nucleotide sequence encoding the polypeptide of the invention is operably linked to additional segments required for transcription of the nucleotide sequence. The vector is typically derived from plasmid or viral DNA. A number of suitable expression vectors for expression in the host cells mentioned herein are commercially available or described in the literature. Useful expression vectors for eukaryotic hosts include, for example, vectors comprising expression control sequences from SV40, bovine papilloma virus, adenovirus and cytomegalovirus. Specific vectors are, e.g.,, pCDNA3.1(+)\Hyg (Invitrogen, Carlsbad, CA, USA) and pCI-neo (Stratagene, La Jolla, CA, USA). Useful expression vectors for yeast cells include the 2μ plasmid and derivatives thereof, the POT1 vector (US 4,931,373), the pJSO37 vector described in Okkels, Ann. New York Acad. Sci. 782, 202-207, 1996, and pPICZ A, B or C (Invitrogen). Useful vectors for insect cells include pVL941, pBG311 (Cate *et al.*, "Isolation of the Bovine and Human Genes for Mullerian Inhibiting Substance And Expression of the Human Gene In Animal Cells", Cell, 45, pp. 685-98 (1986), pBluebac 4.5 and pMelbac (both available from Invitrogen). Useful expression vectors for bacterial hosts include known bacterial plasmids, such as plasmids from *E. coli*, including pBR322, pET3a and pET12a (both from Novagen Inc., WI, USA), wider host range plasmids, such as RP4, phage DNAs, e.g., the numerous derivatives of phage lambda, e.g., NM989, and other DNA phages, such as M13 and filamentous single stranded DNA phages.

Other vectors for use in this invention include those that allow the nucleotide sequence encoding the polypeptide to be amplified in copy number. Such amplifiable vectors are well known in the art. They include, for example, vectors able to be amplified by DHFR amplification (see, e.g.,, Kaufman, U.S. Pat. No. 4,470,461, Kaufman and Sharp, "Construction Of A Modular Dihydrafolate Reductase cDNA Gene: Analysis Of Signals Utilized For Efficient Expression", Mol. Cell. Biol., 2, pp. 1304-19 (1982)) and glutamine synthetase ("GS") amplification (see, e.g.,, US 5,122,464 and EP 338,841).

The recombinant vector may further comprise a DNA sequence enabling the vector to replicate in the host cell in question. An example of such a sequence (when the host cell is a mammalian cell) is the SV40 origin of replication. When the host cell is a yeast cell, suitable sequences enabling the vector to replicate are the yeast plasmid 2μ replication genes REP 1-3 and origin of replication.

The vector may also comprise a selectable marker, e.g., a gene whose product complements a defect in the host cell, such as the gene coding for dihydrofolate reductase (DHFR) or the Schizosaccharomyces pombe TPI gene (described by P.R. Russell, Gene 40, 1985, pp. 125-130), or one which confers resistance to a drug, e.g., ampicillin, kanamycin, tetracyclin, chloramphenicol, neomycin, hygromycin or methotrexate. For *Saccharomyces cerevisiae*, selectable markers include *ura3 and leu2*. For filamentous fungi, selectable markers include *amdS*, *pyrG*, *arcB*, *niaD* and *sC*.

The term "control sequences" is defined herein to include all components which are necessary or advantageous for the expression of the polypeptide of the invention. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader sequence, polyadenylation sequence, propeptide sequence, promoter, enhancer or upstream activating sequence, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter.

A wide variety of expression control sequences may be used in the present invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors as well as any sequence known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof.

Examples of suitable control sequences for directing transcription in mammalian cells include the early and late promoters of SV40 and adenovirus, e.g., the adenovirus 2 major late promoter, the MT-1 (metallothionein gene) promoter, the human cytomegalovirus immediate-early gene promoter (CMV), the human elongation factor 1α (EF-1α) promoter, the *Drosophila* minimal heat shock protein 70 promoter, the Rous Sarcoma Virus (RSV) promoter, the human ubiquitin C (UbC) promoter, the human growth hormone terminator, SV40 or adenovirus Elb region polyadenylation signals and the Kozak consensus sequence (Kozak, M. *J Mol Biol* 1987 Aug 20;196(4):947-50).

In order to improve expression in mammalian cells a synthetic intron may be inserted in the 5' untranslated region of the nucleotide sequence encoding the polypeptide. An example of a synthetic intron is the synthetic intron from the plasmid pCI-Neo (available from Promega Corporation, WI, USA).

Examples of suitable control sequences for directing transcription in insect cells include the polyhedrin promoter, the P10 promoter, the *Autographa californica* polyhedrosis virus basic protein promoter, the baculovirus immediate early gene 1 promoter, the baculovirus 39K delayed-early gene promoter, and the SV40 polyadenylation sequence. Examples of suitable control sequences for use in yeast host cells include the promoters of the yeast α-mating system, the yeast triose phosphate isomerase (TPI) promoter, promoters from yeast glycolytic genes or alcohol dehydrogenase genes, the ADH2-4c promoter, and the inducible GAL promoter. Examples of suitable control sequences for use in filamentous fungal host cells include the ADH3 promoter and terminator, a promoter derived from the genes encoding *Aspergillus oryzae* TAKA amylase-triose phosphate isomerase or alkaline protease, an *A. niger* α-amylase, *A. niger* or *A. nidulans* glucoamylase, *A. nidulans* acetamidase, *Rhizomucor miehei* aspartic proteinase or lipase, the TPI1 terminator and the ADH3 terminator. Examples of suitable control sequences for use in bacterial host cells include promoters of the *lac* system, the *trp* system, the *TAC* or *TRC* system, and the major promoter regions of phage lambda.

The presence or absence of a signal peptide will, e.g.,, depend on the expression host cell used for the production of the polypeptide to be expressed (whether it is an intracellular or extracellular polypeptide) and whether it is desirable to obtain secretion. For use in filamentous fungi, the signal peptide may conveniently be derived from a gene encoding an *Aspergillus* sp. amylase or glucoamylase, a gene encoding a *Rhizomucor miehei* lipase or protease or a *Humicola lanuginosa* lipase. The signal peptide is preferably derived from a gene encoding *A. oryzae* TAKA amylase, *A. niger* neutral α-amylase, *A. niger* acid-stable amylase, or *A. niger* glucoamylase. For use in insect cells, the signal peptide may conveniently be derived from an insect gene (cf. WO 90/05783), such as the *Lepidopteran manduca sexta* adipokinetic hormone precursor, (cf. US 5,023,328), the honeybee melittin (Invitrogen), ecdysteroid UDPglucosyltransferase (egt) (Murphy *et al.*, Protein Expression and Purification 4, 349-357 (1993) or human pancreatic lipase (hpl) (Methods in Enzymology 284, pp. 262-272, 1997). A preferred signal peptide for use in mammalian cells is that of hG-CSF or the murine Ig kappa light chain signal peptide (Coloma, M (1992) J. Imm. Methods 152:89-104). For use in yeast cells suitable signal peptides have been found to be the α-factor signal peptide from *S. cereviciae* (cf. US 4,870,008), a modified carboxypeptidase signal peptide (cf. L.A. Valls *et al.*, Cell 48, 1987, pp. 887-897), the yeast BAR1 signal peptide (cf. WO 87/02670), the yeast aspartic

protease 3 (YAP3) signal peptide (cf. M. Egel-Mitani *et al.*, Yeast 6, 1990, pp. 127-137), and the synthetic leader sequence TA57 (WO98/32867). For use in *E. coli* cells a suitable signal peptide has been found to be the signal peptide *ompA*.

The nucleotide sequence of the invention encoding a polypeptide exhibiting G-CSF activity, whether prepared by site-directed mutagenesis, synthesis, PCR or other methods, may optionally also include a nucleotide sequence that encodes a signal peptide. The signal peptide is present when the polypeptide is to be secreted from the cells in which it is expressed. Such signal peptide, if present, should be one recognized by the cell chosen for expression of the polypeptide. The signal peptide may be homologous (e.g., be that normally associated with hG-CSF) or heterologous (i.e., originating from another source than hG-CSF) to the polypeptide or may be homologous or heterologous to the host cell, i.e., be a signal peptide normally expressed from the host cell or one which is not normally expressed from the host cell. Accordingly, the signal peptide may be prokaryotic, e.g., derived from a bacterium such as *E. coli,* or eukaryotic, e.g., derived from a mammalian, or insect or yeast cell.

Any suitable host may be used to produce the polypeptide or polypeptide part of the conjugate of the invention, including bacteria, fungi (including yeasts), plant, insect, mammal, or other appropriate animal cells or cell lines, as well as transgenic animals or plants. Examples of bacterial host cells include gram-positive bacteria such as strains of *Bacillus,* e.g., *B. brevis* or *B. subtilis, Pseudomonas* or *Streptomyces*, or gram-negative bacteria, such as strains of *E. coli.* The introduction of a vector into a bacterial host cell may, for instance, be effected by protoplast transformation (see, *e.g.*,, Chang and Cohen, 1979, *Molecular General Genetics* 168: 111-115), using competent cells (see, *e.g.*,, Young and Spizizin, 1961, *Journal of Bacteriology* 81: 823-829, or Dubnau and Davidoff-Abelson, 1971, *Journal of Molecular Biology* 56: 209-221), electroporation (see, *e.g.*,, Shigekawa and Dower, 1988, *Biotechniques* 6: 742-751), or conjugation (see, *e.g.*,, Koehler and Thorne, 1987, *Journal of Bacteriology* 169: 5771-5278). Examples of suitable filamentous fungal host cells include strains of *Aspergillus,* e.g., *A. oryzae, A. niger,* or *A. nidulans, Fusarium* or *Trichoderma.* Fungal cells may be transformed by a process involving protoplast formation, transformation of the protoplasts, and regeneration of the cell wall in a manner known *per se*. Suitable procedures for transformation of *Aspergillus* host cells are described in EP 238 023 and US 5,679,543. Suitable methods for transforming *Fusarium* species are described by Malardier *et al.*, 1989, *Gene* 78: 147-156 and WO 96/00787.

Examples of suitable yeast host cells include strains of *Saccharomyces*, e.g., *S. cerevisiae*, *Schizosaccharomyces, Klyveromyces, Pichia,* such as *P. pastoris* or *P. methanolica, Hansenula,* such as *H. polymorpha* or *Yarrowia.* Yeast may be transformed using the procedures described by Becker and Guarente, *In* Abelson, J.N. and Simon, M.I., editors, *Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology*, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito *et al.*, 1983, *Journal of Bacteriology* 153: 163; Hinnen *et al.*, 1978, *Proceedings of the National Academy of Sciences USA* 75: 1920: and as disclosed by Clontech Laboratories, Inc, Palo Alto, CA, USA (in the product protocol for the Yeastmaker™ Yeast Transformation System Kit). Examples of suitable insect host cells include a *Lepidoptora* cell line, such as *Spodoptera frugiperda* (Sf9 or Sf21) or *Trichoplusioa ni* cells (High Five) (US 5,077,214). Transformation of insect cells and production of heterologous polypeptides therein may be performed as described by Invitrogen. Examples of suitable mammalian host cells include Chinese hamster ovary (CHO) cell lines, (e.g., CHO-K1; ATCC CCL-61), Green Monkey cell lines (COS) (e.g., COS 1 (ATCC CRL-1650), COS 7 (ATCC CRL-1651)); mouse cells (e.g., NS/O), Baby Hamster Kidney (BHK) cell lines (e.g., ATCC CRL-1632 or ATCC CCL-10), and human cells (e.g., HEK 293 (ATCC CRL-1573)), as well as plant cells in tissue culture. Additional suitable cell lines are known in the art and available from public depositories such as the American Type Culture Collection, Rockville, Maryland. Methods for introducing exogeneous DNA into mammalian host cells include calcium phosphate-mediated transfection, electroporation, DEAE-dextran mediated transfection, liposome-mediated transfection, viral vectors and the transfection method described by Life Technologies Ltd, Paisley, UK using Lipofectamin 2000. These methods are well known in the art and e.g., described by Ausbel *et al.* (eds.), 1996, Current Protocols in Molecular Biology, John Wiley & Sons, New York, USA. The cultivation of mammalian cells is conducted according to established methods, e.g., as disclosed in (Animal Cell Biotechnology, Methods and Protocols, Edited by Nigel Jenkins, 1999, Human Press Inc, Totowa, New Jersey, USA and Harrison MA and Rae IF, General Techniques of Cell Culture, Cambridge University Press 1997).

In the production methods of the present invention, the cells are cultivated in a nutrient medium suitable for production of the polypeptide using methods known in the art. For example, the cell may be cultivated by shake flask cultivation, small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory

or industrial fermenters performed in a suitable medium and under conditions allowing the polypeptide to be expressed and/or isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable media are available from commercial suppliers or may be prepared according to published compositions (e.g.,, in catalogues of the American Type Culture Collection). If the polypeptide is secreted into the nutrient medium, the polypeptide can be recovered directly from the medium. If the polypeptide is not secreted, it can be recovered from cell lysates.

The resulting polypeptide may be recovered by methods known in the art. For example, the polypeptide may be recovered from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray drying, evaporation, or precipitation.

The polypeptides may be purified by a variety of procedures known in the art including, but not limited to, chromatography (e.g.,, ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (e.g.,, preparative isoelectric focusing), differential solubility (e.g.,, ammonium sulfate precipitation), SDS-PAGE, or extraction (see, e.g.,, *Protein Purification*, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989). Specific methods for purifying polypeptides exhibiting G-CSF activity are described by D. Metcalf and N. A. Nicola in *The hemopoietic colony-stimulating factors*, p. 50-51, Cambridge University Press (1995), by C. S. Bae *et al.*, Appl. Microbiol. Biotechnol, 52:338-344 (1999) and in US 4,810,643.

Pharmaceutical composition of the invention and its use

In a further aspect, the present invention comprises a composition comprising a polypeptide or conjugate as described herein and at least one pharmaceutically acceptable carrier or excipient.

The polypeptide, the conjugate or the pharmaceutical composition according to the invention may be used for the manufacture of a medicament for treatment of diseases, in particular prevention of infection in cancer patients undergoing certain types of chemotherapy, radiation therapy and bone marrow transplantations, mobilisation of progenitor cells for collection in peripheral blood progenitor cell transplantations, treatment of severe chronic or relative leukopenia, treatment of patients with acute myeloid leukaemia, treatment of AIDS or

other immunodeficiency diseases, and for antifungal therapy, in particular for treatment of systemic or invasive candidiasis.

In another aspect the polypeptide, the conjugate or the pharmaceutical composition according to the invention is used in a method of treating a mammal having a general haematopoietic disorder, including those arising from radiation therapy or from chemotherapy, in particular leukopenia, AIDS or other immunodeficiency diseases, comprising administering to a mammal in need thereof such a polypeptide, conjugate or pharmaceutical composition.

The polypeptides and conjugates of the invention will be administered to patients in a "therapeutically effective" dose, i.e., a dose that is sufficient to produced the desired effects in relation to the condition for which it is administered. The exact dose will depend on the disorder to be treated, and will be ascertainable by one skilled in the art using known techniques. The polypeptides or conjugates of the invention may e.g., be administered at a dose similar to that employed in therapy with rhG-CSF such as Neupogen®. A suitable dose of a conjugate of the invention is contemplated to be in the range of about 5-300 microgram/kg body weight (based on the weight of the protein part of the conjugate), e.g., 10-200 microgram/kg, such as 25-100 microgram/kg. It will be apparent to those of skill in the art that an effective amount of a polypeptide, conjugate or composition of the invention depends, inter alia, upon the disease, the dose, the administration schedule, whether the polypeptide or conjugate or composition is administered alone or in conjunction with other therapeutic agents, the serum half-life of the compositions, the general health of the patient, and the frequency of administration. Preferably, the polypeptide, conjugate, preparation or composition of the invention is administered in an effective dose, in particular a dose which is sufficient to normalize the number of leukocytes, in particular neutrophils, in the patient in question. Normalization of the number of leukocytes may be determined by simply counting the number of leukocytes at regular intervals in accordance with established practice.

The polypeptide or conjugate of the invention is preferably administered in a composition including one or more pharmaceutically acceptable carriers or excipients. The polypeptide or conjugate can be formulated into pharmaceutical compositions in a manner known *per se* in the art to result in a polypeptide pharmaceutical that is sufficiently storage-stable and is suitable for administration to humans or animals. The pharmaceutical composition may be formulated in a variety of forms, including as a liquid or gel, or lyophilized, or any other

suitable form. The preferred form will depend upon the particular indication being treated and will be apparent to one of skill in the art.

Accordingly, this invention provides compositions and methods for treating various forms of leukopenia. In particular the polypeptide, conjugate or composition of the invention may be used to prevent infection in cancer patients undergoing certain types of radiation therapy chemotherapy and bone marrow transplantations, to mobilize progenitor cells for collection in peripheral blood progenitor cell transplantations, for treatment of severe chronic or relative leukopenia and to support treatment of patients with acute myeloid leukaemia. Additionally, the polypeptide, conjugate or composition of the invention may be used for treatment of AIDS or other immunodeficiency diseases and for antifungal therapy, in particular for treament of systemic or invasive candidiasis, and for the treatment of bacterial infections.

Drug form

The polypeptide or conjugate of the invention can be used "as is" and/or in a salt form thereof. Suitable salts include, but are not limited to, salts with alkali metals or alkaline earth metals, such as sodium, potassium, calcium and magnesium, as well as e.g., zinc salts. These salts or complexes may by present as a crystalline and/or amorphous structure.

Excipients

"Pharmaceutically acceptable" means a carrier or excipient that at the dosages and concentrations employed does not cause any untoward effects in the patients to whom it is administered. Such pharmaceutically acceptable carriers and excipients are well known in the art (see Remington's Pharmaceutical Sciences, 18th edition, A. R. Gennaro, Ed., Mack Publishing Company [1990]; Pharmaceutical Formulation Development of Peptides and Proteins, S. Frokjaer and L. Hovgaard, Eds., Taylor & Francis [2000] ; and Handbook of Pharmaceutical Excipients, 3rd edition, A. Kibbe, Ed., Pharmaceutical Press [2000]).

Mix of drugs

The pharmaceutical composition of the invention may be administered alone or in conjunction with other therapeutic agents. These agents may be incorporated as part of the same pharmaceutical composition or may be administered separately from the polypeptide or conjugate of the invention, either concurrently or in accordance with another treatment schedule. In addition, the polypeptide, conjugate or pharmaceutical composition of the invention may be used as an adjuvant to other therapies.

Patients

A "patient" for the purposes of the present invention includes both humans and other mammals. Thus the methods are applicable to both human therapy and veterinary applications.

Administration route

The administration of the formulations of the present invention can be performed in a variety of ways, including, but not limited to, orally, subcutaneously, intravenously, intracerebrally, intranasally, transdermally, intraperitoneally, intramuscularly, intrapulmonary, vaginally, rectally, intraocularly, or in any other acceptable manner. The formulations can be administered continuously by infusion, although bolus injection is acceptable, using techniques well known in the art.

Parenterals

An example of a pharmaceutical composition is a solution designed for parenteral administration. Although in many cases pharmaceutical solution formulations are provided in liquid form, appropriate for immediate use, such parenteral formulations may also be provided in frozen or in lyophilized form. In the former case, the composition must be thawed prior to use. The latter form is often used to enhance the stability of the active compound contained in the composition under a wider variety of storage conditions, as it is recognized by those skilled in the art that lyophilized preparations are generally more stable than their liquid counterparts. Such lyophilized preparations are reconstituted prior to use by the addition of one or more suitable pharmaceutically acceptable diluents such as sterile water for injection or sterile physiological saline solution.

In case of parenterals, they are prepared for storage as lyophilized formulations or aqueous solutions by mixing, as appropriate, the polypeptide having the desired degree of purity with one or more pharmaceutically acceptable carriers, excipients or stabilizers typically employed in the art (all of which are termed "excipients"), for example buffering agents, stabilizing agents, preservatives, isotonifiers, non-ionic detergents, antioxidants and/or other miscellaneous additives.

Buffering agents help to maintain the pH in the range which approximates physiological conditions. They are typically present at a concentration ranging from about 2 mM to about 50 mM Suitable buffering agents for use with the present invention include both organic

and inorganic acids and salts thereof such as citrate buffers (e.g.,, monosodium citrate-disodium citrate mixture, citric acid-trisodium citrate mixture, citric acid-monosodium citrate mixture, etc.), succinate buffers (e.g.,, succinic acid-monosodium succinate mixture, succinic acid-sodium hydroxide mixture, succinic acid-disodium succinate mixture, etc.), tartrate buffers (e.g.,, tartaric acid-sodium tartrate mixture, tartaric acid-potassium tartrate mixture, tartaric acid-sodium hydroxide mixture, etc.), fumarate buffers (e.g.,, fumaric acid-monosodium fumarate mixture, fumaric acid-disodium fumarate mixture, monosodium fumarate-disodium fumarate mixture, etc.), gluconate buffers (e.g.,, gluconic acid-sodium glyconate mixture, gluconic acid-sodium hydroxide mixture, gluconic acid-potassium glyuconate mixture, etc.), oxalate buffer (e.g.,, oxalic acid-sodium oxalate mixture, oxalic acid-sodium hydroxide mixture, oxalic acid-potassium oxalate mixture, etc.), lactate buffers (e.g.,, lactic acid-sodium lactate mixture, lactic acid-sodium hydroxide mixture, lactic acid-potassium lactate mixture, etc.) and acetate buffers (e.g.,, acetic acid-sodium acetate mixture, acetic acid-sodium hydroxide mixture, etc.). Additional possibilities are phosphate buffers, histidine buffers and trimethylamine salts such as Tris.

Preservatives are added to retard microbial growth, and are typically added in amounts of about 0.2%-1% (w/v). Suitable preservatives for use with the present invention include phenol, benzyl alcohol, meta-cresol, methyl paraben, propyl paraben, octadecyldimethylbenzyl ammonium chloride, benzalkonium halides (e.g., benzalkonium chloride, bromide or iodide), hexamethonium chloride, alkyl parabens such as methyl or propyl paraben, catechol, resorcinol, cyclohexanol and 3-pentanol.

Isotonicifiers are added to ensure isotonicity of liquid compositions and include polyhydric sugar alcohols, preferably trihydric or higher sugar alcohols, such as glycerin, erythritol, arabitol, xylitol, sorbitol and mannitol. Polyhydric alcohols can be present in an amount between 0.1% and 25% by weight, typically 1% to 5%, taking into account the relative amounts of the other ingredients.

Stabilizers refer to a broad category of excipients which can range in function from a bulking agent to an additive which solubilizes the therapeutic agent or helps to prevent denaturation or adherence to the container wall. Typical stabilizers can be polyhydric sugar alcohols (enumerated above); amino acids such as arginine, lysine, glycine, glutamine, asparagine, histidine, alanine, omithine, L-leucine, 2-phenylalanine, glutamic acid, threonine,

etc., organic sugars or sugar alcohols, such as lactose, trehalose, stachyose, mannitol, sorbitol, xylitol, ribitol, myoinisitol, galactitol, glycerol and the like, including cyclitols such as inositol; polyethylene glycol; amino acid polymers; sulfur-containing reducing agents, such as urea, glutathione, thioctic acid, sodium thioglycolate, thioglycerol, α-monothioglycerol and sodium thiosulfate; low molecular weight polypeptides (i.e., <10 residues); proteins such as human serum albumin, bovine serum albumin, gelatin or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; monosaccharides such as xylose, mannose, fructose and glucose; disaccharides such as lactose, maltose and sucrose; trisaccharides such as raffinose, and polysaccharides such as dextran. Stabilizers are typically present in the range of from 0.1 to 10,000 parts by weight based on the active protein weight.

Non-ionic surfactants or detergents (also known as "wetting agents") may be present to help solubilize the therapeutic agent as well as to protect the therapeutic polypeptide against agitation-induced aggregation, which also permits the formulation to be exposed to shear surface stress without causing denaturation of the polypeptide. Suitable non-ionic surfactants include polysorbates (20, 80, etc.), polyoxamers (184, 188 etc.), Pluronic® polyols, polyoxyethylene sorbitan monoethers (Tween®-20, Tween®-80, etc.).

Additional miscellaneous excipients include bulking agents or fillers (e.g., starch), chelating agents (e.g., EDTA), antioxidants (e.g.,, ascorbic acid, methionine, vitamin E) and cosolvents.

The active ingredient may also be entrapped in microcapsules prepared, for example, by coascervation techniques or by interfacial polymerization, for example hydroxymethylcellulose, gelatin or poly-(methylmethacylate) microcapsules, in colloidal drug delivery systems (for example liposomes, albumin microspheres, microemulsions, nano-particles and nanocapsules) or in macroemulsions. Such techniques are disclosed in Remington's Pharmaceutical Sciences, *supra*.

Parenteral formulations to be used for *in vivo* administration must be sterile. This is readily accomplished, for example, by filtration through sterile filtration membranes.

Sustained release preparations

Suitable examples of sustained-release preparations include semi-permeable matrices of solid hydrophobic polymers containing the polypeptide or conjugate, the matrices having a suitable form such as a film or microcapsules. Examples of sustained-release matrices

include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate) or poly(vinylalcohol)), polylactides, copolymers of L-glutamic acid and ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the ProLease® technology or Lupron Depot® (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for long periods such as up to or over 100 days, certain hydrogels release proteins for shorter time periods. When encapsulated polypeptides remain in the body for a long time, they may denature or aggregate as a result of exposure to moisture at 37°C, resulting in a loss of biological activity and possible changes in immunogenicity. Rational strategies can be devised for stabilization depending on the mechanism involved. For example, if the aggregation mechanism is discovered to be intermolecular S-S bond formation through thio-disulfide interchange, stabilization may be achieved by modifying sulfhydryl residues, lyophilizing from acidic solutions, controlling moisture content, using appropriate additives, and developing specific polymer matrix compositions.

Pulmonary delivery

Formulations suitable for use with a nebulizer, either jet or ultrasonic, will typically comprise the polypeptide or conjugate dissolved in water at a concentration of, e.g.,, about 0.01 to 25 mg of conjugate per mL of solution, preferably about 0.1 to 10 mg/mL. The formulation may also include a buffer and a simple sugar (e.g.,, for protein stabilization and regulation of osmotic pressure), and/or human serum albumin ranging in concentration from 0.1 to 10 mg/ml. Examples of buffers that may be used are sodium acetate, citrate and glycine. Preferably, the buffer will have a composition and molarity suitable to adjust the solution to a pH in the range of 3 to 9. Generally, buffer molarities of from 1 mM to 50 mM are suitable for this purpose. Examples of sugars which can be utilized are lactose, maltose, mannitol, sorbitol, trehalose, and xylose, usually in amounts ranging from 1% to 10% by weight of the formulation.

The nebulizer formulation may also contain a surfactant to reduce or prevent surface induced aggregation of the protein caused by atomization of the solution in forming the aerosol. Various conventional surfactants can be employed, such as polyoxyethylene fatty acid esters and alcohols, and polyoxyethylene sorbitan fatty acid esters. Amounts will generally range

between 0.001% and 4% by weight of the formulation. An especially preferred surfactant for purposes of this invention is polyoxyethylene sorbitan monooleate.

Specific formulations and methods of generating suitable dispersions of liquid particles of the invention are described in WO 94/20069, US 5,915,378, US 5,960,792, US 5,957,124, US 5,934,272, US 5,915,378, US 5,855,564, US 5,826,570 and US 5,522,385 which are hereby incorporated by reference.

Formulations for use with a metered dose inhaler device will generally comprise a finely divided powder. This powder may be produced by lyophilizing and then milling a liquid conjugate formulation and may also contain a stabilizer such as human serum albumin (HSA). Typically, more than 0.5% (w/w) HSA is added. Additionally, one or more sugars or sugar alcohols may be added to the preparation if necessary. Examples include lactose maltose, mannitol, sorbitol, sorbitose, trehalose, xylitol, and xylose. The amount added to the formulation can range from about 0.01 to 200% (w/w), preferably from approximately 1 to 50%, of the conjugate present. Such formulations are then lyophilized and milled to the desired particle size.

The properly sized particles are then suspended in a propellant with the aid of a surfactant. The propellant may be any conventional material employed for this purpose, such as a chlorofluorocarbon, a hydrochlorofluorocarbon, a hydrofluorocarbon, or a hydrocarbon, including trichlorofluoromethane, dichlorodifluoromethane, dichlorotetrafluoroethanol, and 1,1,1,2-tetrafluoroethane, or combinations thereof. Suitable surfactants include sorbitan trioleate and soya lecithin. Oleic acid may also be useful as a surfactant. This mixture is then loaded into the delivery device. An example of a commercially available metered dose inhaler suitable for use in the present invention is the Ventolin metered dose inhaler, manufactured by Glaxo Inc., Research Triangle Park, N.C.

Formulations for powder inhalers will comprise a finely divided dry powder containing conjugate and may also include a bulking agent, such as lactose, sorbitol, sucrose, or mannitol in amounts which facilitate dispersal of the powder from the device, e.g.,, 50% to 90% by weight of the formulation. The particles of the powder shall have aerodynamic properties in the lung corresponding to particles with a density of about 1 g/cm^2 having a median diameter less than 10 micrometers, preferably between 0.5 and 5 micrometers, most preferably of between 1.5 and 3.5 micrometers. An example of a powder inhaler suitable for use in accordance with the teachings herein is the Spinhaler powder inhaler, manufactured by Fisons Corp., Bedford, Mass.

The powders for these devices may be generated and/or delivered by methods disclosed in US 5,997,848, US 5,993,783, US 5,985,248, US 5,976574, US 5,922,354, US 5,785,049 and US 5,654,007.

Mechanical devices designed for pulmonary delivery of therapeutic products, include but are not limited to nebulizers, metered dose inhalers, and powder inhalers, all of which are familiar to those of skill in the art. Specific examples of commercially available devices suitable for the practice of this invention are the Ultravent nebulizer, manufactured by Mallinckrodt, Inc., St. Louis, Missouri; the Acorn II nebulizer, manufactured by Marquest Medical Products, Englewood, Colorado; the Ventolin metered dose inhaler, manufactured by Glaxo Inc., Research Triangle Park, North Carolina; the Spinhaler powder inhaler, manufactured by Fisons Corp., Bedford, Massachusetts; the "standing cloud" device of Inhale Therapeutic Systems, Inc., San Carlos, California; the AIR inhaler manufactured by Alkermes, Cambridge, Massachusetts; and the AERx pulmonary drug delivery system manufactured by Aradigm Corporation, Hayward, California.

Also contemplated is use of a nucleotide sequence encoding a polypeptide of the invention in gene therapy applications. In particular, it may be of interest to use a nucleotide sequence encoding a polypeptide as described in the section above entitled "Conjugate of the invention wherein the non-polypeptide moiety is a carbohydrate moiety". The glycosylation of the polypeptides is thus achieved during the course of the gene therapy, i.e., after expression of the nucleotide sequence in the human body.

Gene therapy applications contemplated include treatment of those diseases in which the polypeptide is expected to provide an effective therapy. These include treatment of various forms of leukopenia in particular prevention of infection in cancer patients undergoing certain types of chemotherapy and bone marrow transplantations, mobilization of progenitor cells for collection in peripheral blood progenitor cell transplantations, treatment of severe chronic or relative leukopenia, treatment of patients with acute myeloid leukaemia, and treatment of AIDS or other immunodeficiency diseases.

Local delivery of G-CSF using gene therapy may provide the therapeutic agent to the target area while avoiding potential toxicity problems associated with non-specific administration.

Both *in vitro* and *in vivo* gene therapy methodologies are contemplated.

Several methods for transferring potentially therapeutic genes to defined cell populations are known. For further reference see, e.g.,, Mulligan, "The Basic Science Of Gene Therapy", Science, 260, pp. 926-31 (1993). These methods include:

Direct gene transfer, e.g.,, as disclosed by Wolff *et al.*, "Direct Gene transfer Into Mouse Muscle In vivo", Science 247, pp. 1465-68 (1990);

Liposome-mediated DNA transfer, e.g.,, as disclosed by Caplen *et al.*, "Liposome-mediated CFTR Gene Transfer to the Nasal Epithelium Of Patients With Cystic Fibrosis" Nature Med., 3, pp. 39-46 (1995); Crystal, "The Gene As A Drug", Nature Med., 1, pp.- 15-17 (1995); Gao and Huang, "A Novel Cationic Liposome Reagent For Efficient Transfection of Mammalian Cells", Biochem.Biophys Res. Comm., 179, pp. 280-85 (1991);

Retrovirus-mediated DNA transfer, e.g.,, as disclosed by Kay *et al.*, "In vivo Gene Therapy of Hemophilia B: Sustained Partial Correction In Factor IX-Deficient Dogs", Science, 262, pp. 117-19 (1993); Anderson, "Human Gene Therapy", Science, 256, pp.808-13(1992);

DNA Virus-mediated DNA transfer. Such DNA viruses include adenoviruses (preferably Ad-2 or Ad-5 based vectors), herpes viruses (preferably herpes simplex virus based vectors), and parvoviruses (preferably "defective" or non-autonomous parvovirus based vectors, more preferably adeno-associated virus based vectors, most preferably AAV-2 based vectors). See, e.g.,, Ali *et al.*, "The Use Of DNA Viruses as Vectors for Gene Therapy", Gene Therapy, 1, pp. 367-84 (1994); US 4,797,368, and US 5,139,941.

MATERIALS AND METHODS

Methods used to determine the amino acids to be modified

Accessible Surface Area (ASA)

A 3D ensemble of 10 structures determined by NMR spectroscopy (Zink et al. (1994) Biochemistry 33: 8453-8463) is available from the Protein Data Bank (PDB) (www.rcsb.org/pdb/). This information can be entered into the computer program Access (B. Lee and F.M. Richards, J. Mol. Biol. 55: 379-400 (1971)) version 2 (© 1983 Yale University) and used to compute the accessible surface area (ASA) of the individual atoms in the structure. This method typically uses a probe size of 1.4Å and defines the Accessible Surface Area (ASA) as the area formed by the centre of the probe. Prior to this calculation all water molecules and all hydrogen atoms should be removed from the coordinate set as should other atoms not directly related to the protein.

Fractional ASA of side chain

The fractional ASA of the side chain atoms is computed by division of the sum of the ASA of the atoms in the side chain with a value representing the ASA of the side chain atoms of that residue type in an extended ALA-x-ALA tripeptide. See Hubbard, Campbell & Thornton (1991) J.Mol.Biol.220, 507-530. For this example the CA atom is regarded as a part of the side chain of glycine residues but not for the remaining residues. The values in the following table are used as standard 100% ASA for the side chain:

Ala	69.23	$Å^2$	Leu	140.76	$Å^2$
Arg	200.35	$Å^2$	Lys	162.50	$Å^2$
Asn	106.25	$Å^2$	Met	156.08	$Å^2$
Asp	102.06	$Å^2$	Phe	163.90	$Å^2$
Cys	96.69	$Å^2$	Pro	119.65	$Å^2$
Gln	140.58	$Å^2$	Ser	78.16	$Å^2$
Glu	134.61	$Å^2$	Thr	101.67	$Å^2$
Gly	32.28	$Å^2$	Trp	210.89	$Å^2$
His	147.00	$Å^2$	Tyr	176.61	$Å^2$
Ile	137.91	$Å^2$	Val	114.14	$Å^2$

Residues not detected in the structure are defined as having 100% exposure as they are thought to reside in flexible regions.

Determining distances between atoms:

The distance between atoms is most easily determined using molecular graphics software, e.g., InsightII® v. 98.0, MSI INC.

General considerations regarding amino acid residues to be modified

As explained above, amino acid residues to be modified in accordance with the present invention are preferably those whose side chains are surface exposed, in particular those with more than about 25% of the side chain exposed at the surface of the molecule, and more preferably those with more than 50% side chain exposure. Another consideration is that residues located in receptor interfaces are preferably excluded so as to avoid or at least minimize possible interference with receptor binding or activation. A further consideration is that residues that are less than 10Å from the nearest Lys (Glu, Asp) CB-CB (CA for Gly) should also be excluded. Finally, preferred positions for modification are in particular those that have a hydrophilic and/or

charged residue, i.e., Asp, Asn, Glu, Gln, Arg, His, Tyr, Ser and Thr, positions that have an arginine residue being especially preferred.

Identifying G-CSF amino acid residues for modification

The information below illustrates the factors that generally should be taken into consideration when identifying amino acid residues to be modified in accordance with the present invention.

Three-dimensional structures have been reported for human G-CSF by X-ray crystallography (Hill et al. (1993) Proc.Natl.Acad.Sci.USA 90: 5167-5171) and by NMR spectroscopy (Zink et al. (1994) Biochemistry 33: 8453-8463). As mentioned above, Aritomi et al.(Nature 401:713-717, 1999) have identified the following hG-CSF residues as being part of the receptor binding interfaces: G4, P5, A6, S7, S8, L9, P10, Q11, S12, L15, K16, E19, Q20, L108, D109, D112, T115, T116, Q119, E122, E123, and L124. Thus, although it is possible to modify these residues, it is preferred that these residues are excluded from modification.

Using the 10 NMR structures of G-CSF identified by Zink et al. (1994) as input structures followed by a computation of the average ASA of the side chain, the following residues have been identified as having more than 25% ASA: M0, T1, P2, L3, G4, P5, A6, S7, S8, L9, P10, Q11, S12, F13, L14, L15, K16, C17, E19, Q20, V21, R22, K23, Q25, G26, D27, A29, A30, E33, K34, C36, A37, T38, Y39, K40, L41, H43, P44, E45, E46, V48, L49, L50, H52, S53, L54, I56, P57, P60, L61, S62, S63, P65, S66, Q67, A68, L69, Q70, L71, A72, G73, C74, S76, Q77, L78, S80, F83, Q86, G87, Q90, E93, G94, S96, P97, E98, L99, G100, P101, T102, D104, T105, Q107, L108, D109, A111, D112, F113, T115, T116, W118, Q119, Q120, M121, E122, E123, L124, M126, A127, P128, A129, L130, Q131, P132, T133, Q134, G135, A136, M137, P138, A139, A141, S142, A143, F144, Q145, R146, R147, S155, H156, Q158, S159, L161, E162, V163, S164, Y165, R166, V167, L168, R169, H170, L171, A172, Q173, P174.

Similarly, the following residues have more than 50% ASA: M0, T1, P2, L3, G4, P5, A6, S7, S8, L9, P10, Q11, S12, F13, L14, L15, K16, C17, E19, Q20, R22, K23, G26, D27, A30, E33, K34, T38, K40, L41, H43, P44, E45, E46, L49, L50, S53, P57, P60, L61, S62, S63, P65, S66, Q67, A68, L69, Q70, L71, A72, G73, S80, F83, Q90, G94, P97, E98, P101, D104, T105, L108, D112, F113, T115, T116, Q119, Q120, E122, E123, L124, M126, P128, A129, L130, Q131, P132, T133, Q134, G135, A136, A139, A141, S142, A143, F144, R147, S155, S159, E162, R166, V167, R169, H170, L171, A172, Q173, P174.

The molecular graphics program InsightII® v.98.0 was used to determine residues having their CB atom (CA in the case of glycine) at a distance of more than 15Å from the nearest amine group, defined as the NZ atoms of lysine and the N atom of the N-terminal residue T1. The following list includes the residues that fulfill this criteria in at least one of the 10 NMR structures. G4, P5, A6, S7, S8, L9, P10, Q11, L14, L15, L18, V21, R22, Q25, G26, D27, G28, A29, Q32, L35, C36, T38, Y39, C42, H43, P44, E45, E46, L47, V48, L49, L50, G51, H52, S53, L54, G55, I56, P57, W58, A59, P60, L61, S62, S63, C64, P65, S66, Q67, A68, L69, Q70, L71, A72, G73, C74, L75, S76, Q77, L78, H79, S80, G81, L82, F83, L84, Y85, Q86, G87, L88, L89, Q90, A91, L92, E93, G94, I95, S96, P97, E98, L99, G100, P101, T102, L103, D104, T105, L106, Q107, L108, D109, V110, A111, D112, F113, A114, T115, T116, I117, W118, Q119, Q120, M121, E122, E123, L124, G125, M126, A127, P128, A129, L130, Q131, P132, T133, Q134, G135, A136, M137, P138, A139, F140, A141, S142, A143, F144, Q145, R146, R147, A148, G149, G150, V151, L152, V153, A154, S155, H156, L157, Q158, S159, F160, L161, E162, V163, S164, Y165, R166, V167, L168, R169, H170, L171, A172, Q173, P174.

The InsightII® v.98.0 program was similarly used to determine residues having their CB atom (CA atom in the case of glycine) at a distance of more than 10Å from the nearest acidic group, defined as the CG atoms of aspartic acid, the CD atoms of glutamic acid and the C atom of the C-terminal residue P174. The following list includes the residues that fulfill this criteria in at least one of the 10 NMR structures. M0, T1, P2, L3, G4, P5, A6, S7, S8, L9, P10, Q11, S12, F13, L14, T38, Y39, K40, L41, C42, L50, G51, H52, S53, L54, G55, I56, P57, W58, A59, P60, L61, S62, S63, C64, P65, S66, Q67, A68, L69, Q70, L71, A72, G73, C74, L75, S76, Q77, L78, H79, S80, G81, L82, F83, L84, Y85, Q86, G87, L88, I117, M126, A127, P128, A129, L130, Q131, P132, T133, Q134, G135, A136, M137, P138, A139, F140, A141, S142, A143, F144, Q145, R146, R147, A148, G149, G150, V151, L152, V153, A154, S155, H156, L157, V167, L168, R169, H170, L171.

By combining and comparing the above lists, it is possible to select individual amino acid residues for modification to result in a list containing a limited number of amino acid residues whose modification in a given G-CSF polypeptide is likely to result in desired properties.

Methods for PEGylation of hG-CSF and variants thereof

PEGylation of hG-CSF and variants thereof in solution

Human G-CSF and variants thereof are PEGylated at a concentration of 250 μg/ml in 50 mM sodium phosphate, 100 mM NaCl, pH 8.5. The molar surplus of PEG is 100 times with respect to PEGylation sites on the protein. The reaction mixture is placed in a thermo mixer for 30 minutes at 37°C at 1200 rpm. After 30 minutes, quenching of the reaction is obtained by adding a molar excess of glycine.

Cation exchange chromatography is applied to remove excess PEG, glycine and other by-products from the reaction mixture. The PEGylation reaction mixture is diluted with 20 mM sodium citrate pH 2.5 until the ionic strength is less than 7 mS/cm. pH is adjusted to 2.5 using 5 N HCl. The mixture is applied to a SP-sepharose FF column equilibrated with 20 mM sodium citrate pH 2.5. Unbound material is washed off the column using 4 column volumes of equilibration buffer. PEGylated protein is eluted in three column volumes by adding 20 mM sodium citrate, 750 mM sodium chloride. Pure PEGylated G-CSF is concentrated and buffer exchange is performed using VivaSpin concentration devices, molecular weight cut-off (mwco): 10kDa.

PEGylation in microtiter plates of a tagged polypeptide with G-CSF activity

A polypeptide exhibiting G-CSF activity is expressed with a suitable tag, e.g., any of the tags exemplified in the general description above, and culture broth is transferred to one or more wells of a microtiter plate capable of immobilising the tagged polypeptide. When the tag is Met-Lys-His-Gln-His-Gln-His-Gln-His-Gln-His-Gln-His-Gln-Gln, a nickel-nitrilotriacetic acid (Ni-NTA) HisSorb microtiter plate commercially available from QIAGEN can be used.

After immobilization of the tagged polypeptide to the microtiter plate, the wells are washed in a buffer suitable for binding and subsequent PEGylation followed by incubating the wells with the activated PEG of choice. As an example, M-SPA-5000 from Shearwater Polymers is used. The molar ratio of activated PEG to polypeptide should be optimized, but will typically be greater than 10:1, e.g., up to about 100:1 or higher. After a suitable reaction time at ambient temperature, typically around 1 hour, the reaction is stopped by removal of the activated PEG solution. The conjugated protein is eluted from the plate by incubation with a suitable buffer. Suitable elution buffers may contain imidazole, excess NTA or another chelating compound. The conjugated protein is assayed for biological activity and immunogenicity as

appropriate. The tag may optionally be cleaved off using a method known in the art, e.g., using diaminopeptidase the Gln in pos –1 can be converted to pyroglutamyl with GCT (glutamylcyclotransferase) and finally cleaved off with PGAP (pyro-glutamyl-aminopeptidase), giving the untagged protein. The process involves several steps of metal chelate affinity chromatography. Alternatively, the tagged polypeptide may be conjugated.

PEGylation of a polypeptide exhibiting hG-CSF activity and having a blocked receptor-binding site

In order to optimize PEGylation of hG-CSF in a manner excluding PEGylation of lysines involved in receptor recognition, the following method has been developed:

Purified hG-CSF is obtained as described in Example 3. A homodimer complex consisting of an hG-CSF polypeptide and the soluble domain of the G-CSF receptor in a 2:2 stochiometry is formed in a phosphate-buffered saline solution (PBS) buffer at pH 7. The concentration of hG-CSF polypeptide is approximately 20 μg/ml or 1 μM and the receptor is present at an equimolar concentration.

M-SPA-5000 from Shearwater Polymers, Inc. is added at 3 different concentration levels corresponding to a 5, 20 and 100 molar excess of hG-CSF polypeptide. The reaction time is 30 min at room temperature. After the 30 min reaction period, the pH of the reaction mixture is adjusted to pH 2.0 and the reaction mixture is applied to a Vydac C18 column and eluted with an acetonitrile gradient essentially as described (Utsumi et al., J. Biochem., Vol. 101, 1199-1208, (1987). Alternatively, an isopropanol gradient can be used.

Fractions are analyzed using the primary screening assay described herein and active PEGylated hG-CSF polypeptide obtained by this method is stored at -80°C in PBS, pH 7 containing 1 mg/ml human serum albumin (HSA).

<u>Methods used to characterize conjugated and non-conjugated hG-CSF and variants thereof</u>

Determination of the molecular size of hG-CSF and variants thereof

The molecular weight of conjugated or non-conjugated hG-CSF or variants thereof is determined by either SDS-PAGE, gel filtration, matrix assisted laser desorption mass spectrometry or equilibrium centrifugation.

Determination of polypeptide concentration

The concentration of a polypeptide can be measured using optical density measurements at 280 nm, an enzyme-linked immunoadsorption assay (ELISA), a radio-immunoassay (RIA), or other such immunodetection techniques well known in the art. Furthermore, the polypeptide concentration in a sample can be measured with the Biacore® instrument using a Biacore® chip coated with an antibody specific for the polypeptide.

Such an antibody can be coupled covalently to the Biacore® chip by various chemistries. Alternatively, the antibody can be bound non-covalently e.g., by means of an antibody specific for the Fc portion of the anti-polypeptide antibody. The Fc specific antibody is first coupled covalently to the chip. The anti-polypeptide antibody is then flowed over the chip and is bound by the first antibody in a directed fashion. Furthermore, biotinylated antibodies can be immobilised using a streptavidin coated surface (e.g., Biacore Sensor Chip SA®) (Real-Time Analysis of Biomolecular Interactions, Nagata and Handa (Eds.), 2000, Springer Verlag, Tokyo; Biacore 2000 Instrument Handbook, 1999, Biacore AB).

When the sample is flowed over the chip the polypeptide will bind to the coated antibody and the increase in mass can be measured. By using a preparation of the polypeptide in a known concentration, a standard curve can be established and subsequently the concentration of the polypeptide in the sample can be determined. After each injection of sample the sensor chip is regenerated by a suitable eluent (e.g., a low pH buffer) that removes the bound analyte.

Generally, the applied antibodies will be monoclonal antibodies raised against the wild type polypeptide. Introduction of mutations or other manipulations of the wild type polypeptide (extra glycosylations or polymer conjugations) may alter the recognition by such antibodies. Furthermore, such manipulations that give rise to an increased molecular weight of the polypeptide will result in an increased plasmon resonance signal. Consequently, it is necessary to establish a standard curve for every molecule to be tested.

Methods used to determine the *in vitro* and *in vivo* activity of conjugated and non-conjugated hG-CSF and variants thereof

Primary assay 1 – in vitro G-CSF activity assay

Proliferation of the murine cell line NFS-60 (obtained from Dr. J. Ihle, St. Jude Children's Research Hospital, Tennessee, USA) is dependent on the presence of active G-SCF in the growth medium. Thus, the *in vitro* biological activity of hG-CSF and variants thereof can be

determined by measuring the number of dividing NFS-60 cells after addition of a G-CSF sample to the growth medium followed by incubation over a fixed period of time.

NFS-60 cells are maintained in Iscoves DME Medium containing 10% w/w FBS (fetal bovine serum), 1% w/w Pen/Strep, 10 µg per litre hG-CSF and 2 mM Glutamax. Prior to sample addition, cells are washed twice in growth medium without hG-CSF and diluted to a concentration of 2.2×10^5 cells per ml. 100 µl of the cell suspension is added to each well of a 96 well microtiter plate (Corning).

Samples containing conjugated or non-conjugated G-CSF or variants thereof are diluted to concentrations between 1.1×10^{-6} M and 1.1×10^{-13} M in the growth medium. 10 µl of each sample is added to 3 wells containing NFS-60 cells. A control consisting of 10 µl of mammalian growth medium is added to 8 wells on each microtiter plate. The cells are incubated for 48 hours (37°C, 5% CO_2) and the number of dividing cells in each well is quantified using the WST-1 cell proliferation agent (Roche Diagnostics GmbH, Mannheim, Germany). 0.01 ml WST-1 is added to the wells followed by incubation for 150 min. at 37°C in a 5% CO_2 air atmosphere. The cleavage of the tetrazolium salt WST-1 by mitochondrial dehydrogenases in viable cells results in the formation of formazan that is quantified by measuring the absorbance at 450 nm. Hereby, the number of viable cells in each well is quantified.

Based on these measurements, dose-response curves for each conjugated and non-conjugated G-CSF molecule or variants thereof are calculated, after which the EC50 value for each molecule can be determined. This value is equal to the amount of active G-CSF protein that is necessary to obtain 50% of the maximum proliferation activity of non-conjugated human G-CSF. Thus, the EC50 value is a direct measurement of the *in vitro* activity of the given protein.

Primary assay 2 – in vitro G-CSF activity assay

The murine hematopoietic cell line BaF3 is transfected with a plasmid carrying the human G-CSF receptor and the promoter of the transcription regulator, *fos*, in front of the luciferase reporter gene. Upon stimulation of such a cell line with a G-CSF sample, a number of intracellular reactions lead to stimulation of *fos* expression, and consequently to expression of luciferase. This stimulation is monitored by the Steady-Glo™ Luciferase Assay System (Promega, Cat. No. E2510) whereby the *in vitro* activity of the G-CSF sample may be quantified.

BaF3/hGCSF-R/pfos-lux cells are maintained at 37°C in a humidified 5% CO_2 atmosphere in complete culture media (RPMI-1640/HEPES (Gibco/BRL, Cat. No. 22400), 10%

FBS (HyClone, characterized), 1x Penicillin/Streptomycin (Gibco/BRL, Cat. No. 15140-122), 1x L-Glutamine (Gibco/BRL, Cat. No. 25030-081), 10% WEHI-3 conditioned media (source of muIL-3), and grown to a density of 5 x 10^5 cells/mL (confluent). The cells are reseeded at about 2 x 10^4 cells/mL every 2-3 days.

One day prior to the assay, log-phase cells are resuspended at 2 x 10^5 cells/mL in starving media (DMEM/F-12 (Gibco/BRL, Cat. No. 11039), 1% BSA (Sigma, Cat. No. A3675), 1x Penicillin/Streptomycin (Gibco/BRL, Cat. No. 15140-122), 1x L-Glutamine (Gibco/BRL, Cat. No. 25030-081), 0.1% WEHI-3 conditioned media) and starved for 20 hours. The cells are washed twice with PBS and tested for viability using Trypan Blue viability staining. The cells are resuspended in assay media (RPMI-1640 (phenol-red free, Gibco/BRL, Cat. No. 11835), 25 mM HEPES, 1% BSA (Sigma, Cat. No. A3675), 1x Penicillin/Streptomycin (Gibco/BRL, Cat. No. 15140-122), 1x L-Glutamine (Gibco/BRL, Cat. No. 25030-081) at 4 x 10^6 cells/mL, and 50 µL are aliquotted into each well of a 96-well microtiter plate (Corning). Samples containing conjugated or non-conjugated G-CSF or variants thereof are diluted to concentrations between $1.1x10^{-7}$ M and $1.1x10^{-12}$ M in the assay medium. 50 µl of each sample is added to 3 wells containing BaF3/hGCSF-R/pfos-lux cells. A negative control consisting of 50 µl of medium is added to 8 wells on each microtiter plate. The plates are mixed gently and incubated for 2 hours at 37°C. The luciferase activity is measured by following the Promega Steady-Glo™ protocol (Promega Steady-Glo™ Luciferase Assay System, Cat. No. E2510). 100 µL of substrate is added per well followed by gentle mixing. Luminescence is measured on a TopCount luminometer (Packard) in SPC (single photon counting) mode.

Based on these measurements, dose-response curves for each conjugated and non-conjugated G-CSF molecule or variants thereof are calculated, after which the EC50 value for each molecule can be determined.

Secondary assay – binding affinity of G-CSF or variants thereof to the hG-CSF receptor

Binding of rhG-CSF or variants thereof to the hG-CSF receptor is studied using standard binding assays. The receptors may be purified extracellular receptor domains, receptors bound to purified cellular plasma membranes, or whole cells - the cellular sources being either cell lines that inherently express G-CSF receptors (e.g., NFS-60) or cells transfected with cDNAs encoding the receptors. The ability of rhG-CSF or variants thereof to compete for the binding sites with native G-CSF is analyzed by incubating with a labeled G-CSF-analog, for instance

biotinylated hG-CSF or radioiodinated hG-CSF. An example of such an assay is described by Yamasaki et al. (Drugs. Exptl. Clin. Res. 24:191-196 (1998)).

The extracellular domains of the hG-CSF receptor can optionally be coupled to Fc and immobilized in 96 well plates. RhG-CSF or variants thereof are subsequently added and the binding of these is detected using either specific anti-hG-CSF antibodies or biotinylated or radioiodinated hG-CSF.

Measurement of the in vivo half-life of conjugated and non-conjugated rhG-CSF and variants thereof

An important aspect of the invention is the prolonged biological half-life that is obtained by construction of a hG-CSF with or without conjugation of the polypeptide to the polymer moiety. The rapid decrease of hG-CSF serum concentrations has made it important to evaluate biological responses to treatment with conjugated and non-conjugated hG-CSF and variants thereof. Preferably, the conjugated and non-conjugated hG-CSF and variants thereof of the present invention have prolonged serum half-lives also after i.v. administration, making it possible to measure by e.g., an ELISA method or by the primary screening assay. Measurement of *in vivo* biological half-life was carried out as described below.

Male Sprague Dawley rats (7 weeks old) were used. On the day of administration, the weights of the animals were measured (280-310 gram per animal). 100 µg per kg body weight of the non-conjugated and conjugated hG-CSF samples were each injected intravenously into the tail vein of three rats. At 1 minute, 30 minutes, 1, 2, 4, 6, and 24 hours after the injection, 500 µl of blood was withdrawn from the eyes of each rat while under CO_2 -anaesthesia. The blood samples were stored at room temperature for 1½ hours followed by isolation of serum by centrifugation (4°C, 18000xg for 5 minutes). The serum samples were stored at –80°C until the day of analysis. The amount of active G-CSF in the serum samples was quantified by the G-CSF *in vitro* activity assays (see primary assay 1 and 2) after thawing the samples on ice.

Another example of an assay for the measurement of *in vivo* half-life of G-CSF or variants thereof is described in US 5,824,778, the content of which is hereby incorporated by reference.

Measurement of the in vivo biological activity of conjugated and non-conjugated hG-CSF and variants thereof

Measurement of the *in vivo* biological effects of hG-CSF in SPF Sprague Dawley rats (purchased from M & B A/S, Denmark) is used to evaluate the biological efficacy of conjugated and non-conjugated G-CSF and variants thereof.

On the day of arrival the rats are randomly allocated into groups of 6. The animals are acclimatised for a period of 7 days wherein individuals in poor condition or at extreme weights are rejected. The weight range of the rats at the start of the acclimatization period is 250-270g.

On the day of administration the rats are fasted for 16 hours followed by subcutaneous injection of 100 µg per kg body weight of hG-CSF or a variant thereof. Each hG-CSF sample is injected into a group of 6 randomized rats. Blood samples of 300 µl EDTA stabilised blood are drawn from a tail vein of the rats prior to dosing and at 6, 12, 24, 36, 48, 72, 96, 120 and 144 hours after dosing. The blood samples are analyzed for the following haematological parameters: Haemoglobin, red blood cell count, haematocrit, mean cell volume, mean cell haemoglobin concentration, mean cell haemoglobin, white blood cell count, differential leucocyte count (neutrophils, lymphocytes, eosinophils, basophils, monocytes). On the basis of these measurements the biological efficacy of conjugated and non-conjugated hG-CSF and variants thereof is evaluated.

Further examples of assays for the measurement of i*n vivo* biological activity of hG-CSF or variants thereof are described in US 5,681,720, US 5,795,968, US 5,824,778, US 5,985,265 and by Bowen et al., Experimental Hematology 27:425-432 (1999).

Determination of polypeptide receptor-binding affinity (on- and off-rate)

The strength of the binding between a receptor and ligand can be measured using an enzyme-linked immunoadsorption assay (ELISA), a radio-immunoassay (RIA), or other such immunodetection techniques well known in the art. The ligand-receptor binding interaction may also be measured with the Biacore® instrument, which exploits plasmon resonance detection (Zhou et al., Biochemistry, 1993, 32, 8193-98; Faegerstram and O'Shannessy, 1993, In Handbook of Affinity Chromatography, 229-52, Marcel Dekker, Inc., NY).

The Biacore® technology allows one to bind receptor to a gold surface and to flow ligand over it. Plasmon resonance detection gives direct quantification of the amount of

mass bound to the surface in real time. This technique yields both on- and off-rate constants and thus a ligand-receptor dissociation constant and an affinity constant can be directly determined.

In vitro immunogenicity test of hG-CSF conjugates

The reduced immunogenicity of a conjugate of the invention can be determined by use of an ELISA method measuring the immunoreactivity of the conjugate relative to a reference molecule or preparation. The reference molecule or preparation is normally a recombinant human G-CSF preparation such as Neupogen® or another recombinant human G-CSF preparation, e.g., an N-terminally PEGylated rhG-CSF molecule as described in US 5,824,784. The ELISA method is based on antibodies from patients treated with one of these recombinant G-CSF-preparations. The immunogenicity is considered to be reduced when the conjugate of the invention has a statistically significant lower response in the assay than the reference molecule or preparation.

Neutralisation of activity in G-CSF bioassay

The neutralisation of hG-CSF conjugates by anti-G-CSF sera is analyzed using the G-CSF bioassay described above.

Sera from patients treated with the G-CSF reference molecule or from immunized animals are used. Sera are added either in a fixed concentration (dilution 1:20-1:500 (pt sera) or 20-1000 ng/ml (animal sera)) or in five-fold serial dilutions of sera starting at 1:20 (pt sera) or 1000 ng/ml (animal sera). HG-CSF conjugate is added either in seven fold-dilutions starting at 10 nM or in a fixed concentration (1-100 pM) in a total volume of 80μl DMEM medium + 10% FCS. The sera are incubated for 1 hr. at 37°C with hG-CSF conjugate.

The samples (0.01 ml) are then transferred to 96 well tissue culture plates containing NFS-60 cells in 0.1 ml DMEM media. The cultures are incubated for 48 hours at 37°C in a 5% CO_2 air atmosphere. 0.01 ml WST-1 (WST-1 cell proliferation agent, Roche Diagnostics GmbH, Mannheim, Germany) is added to the cultures and incubated for 150 min. at 37°C in a 5% CO_2 air atmosphere. The cleavage of the tetrazolium salt WST-1 by mitochondrial dehydrogenases in viable cells results in the formation of formazan that is quantified by measuring the absorbance at 450 nm.

When hG-CSF conjugate samples are titrated in the presence of a fixed amount of serum, the neutralising effect is defined as fold inhibition (FI) quantified as EC50(with

serum)/EC50(without serum). The reduction of antibody neutralisation of G-CSF variant proteins is defined as

$$\left(1 - \frac{(\text{FI variant} - 1)}{(\text{FI wt} - 1)}\right) \times 100\%$$

EXAMPLES

EXAMPLE 1. CONSTRUCTION AND CLONING OF SYNTHETIC GENES ENCODING HG-CSF

The following DNA fragments were synthesised following the general procedure described by Stemmer *et al.* (1995), *Gene* 164, pp. 49-53:

Fragment 1, consisting of a *Bam HI* digestion site, a sequence encoding the *YAP3* signal peptide (WO 98/32867), a sequence encoding the TA57 leader sequence (WO 98/32867), a sequence encoding a KEX2 protease recognition site (AAAAGA), a sequence encoding hG-CSF with its codon usage optimised for expression in *E. coli,* (SEQ ID NO:2) and a *Xba I* digestion site.

Fragment 2, consisting of a *Bam HI* digestion site, a sequence encoding the *YAP3* signal peptide (WO 98/32867), a sequence encoding the TA57 leader sequence (WO 98/32867), a sequence encoding a histidine tag (SEQ ID NO:5), a sequence encoding a KEX2 protease recognition site (AAAAGA), a sequence encoding hG-CSF with its codon usage optimised for expression in *E. coli,* (SEQ ID NO:2) and a *Xba I* digestion site.

Fragment 3, consisting of a *Nde I* digestion site, a sequence encoding the *OmpA* signal peptide (SEQ ID NO:3), a sequence encoding hG-CSF with its codon usage optimised for expression in *E. coli,* (SEQ ID NO:2) and a *Bam HI* digestion site.

Fragment 4, consisting of a *Bam HI* digestion site, the Kozak consensus sequence (Kozak, M. *J Mol Biol* 1987 Aug 20;196(4):947-50), a sequence encoding the hG-CSF signal peptide (SEQ ID NO:7) and hG-CSF with its codon usage optimised for expression in CHO cells (SEQ ID NO:8) and a *Xba I* digestion site.

DNA fragment 1 and 2 were inserted into the *Bam* HI and *Xba* I digestion sites in plasmid pJSO37 (Okkels, Ann. New York Acad. Sci. 782:202-207, 1996) using standard DNA techniques. This resulted in plasmids pG-CSFcerevisiae and pHISG-CSFcerevisiae.

DNA fragment 3 was inserted into the *Nde* I and *Bam HI* digestion sites in plasmid pET12a (Invitrogen) using standard DNA techniques. This resulted in plasmid pG-CSFcoli

DNA fragment 4 was inserted into the *Bam* HI and *Xba* I digestion sites in plasmid pcDNA3.1(+) (Invitrogen) using standard DNA techniques. This resulted in plasmid pG-CSFCHO.

EXAMPLE 2. EXPRESSION OF HG-CSF IN S. CEREVISIAE AND E. COLI

Transformation of *Saccharomyces cerevisiae* YNG318 (available from the American Type Culture Collection, VA, USA as ATCC 208973) with either plasmid pG-CSFcerevisiae or pHISG-CSFcerevisiae, isolation of transformants containing either of the two plasmids, and subsequent extracellular expression of hG-CSF without and with the HIS tag, respectively, was performed using standard techniques described in the literature. Transformation of *E. coli* BL21 (DE3) (Novagen, Cat. No. 69387-3) with pG-CSFcoli, isolation of transformants containing the plasmid and subsequent expression of hG-CSF in the supernatant and in the periplasm of the cell was performed as described in the pET System Manual (8th edition) from Novagen.

Expression of hG-CSF by *S. cerevisiae* and *E. coli* was verified by Western Blot analysis using the ImmunoPure Ultra-Sensitive ABC Rabbit IgG Staining kit (Pierce) and a polyclonal antibody against hG-CSF (Pepro Tech EC Ltd.). It was observed that the protein had the correct size.

The expression levels of hG-CSF with and without the N-terminal histidine tag in *S. cerevisiae* and *E. coli* were quantified using a commercially available G-CSF specific ELISA kit (Quantikine Human G-CSF Immunoassay, R&D Systems Cat. No. DCS50). The measured values are listed below.

Expression system	Expression level (mg G-CSF per L)
hG-CSF in *S. cerevisiae*	30
hG-CSF with histidine tag in *S. cerevisiae*	25
hG-CSF in *E. coli*	0.05

EXAMPLE 3. PURIFICATION OF HG-CSF AND VARIANTS THEREOF FROM S. CEREVISIAE CULTURE SUPERNATANTS

Purification of hG-CSF was performed as follows:

Cells are removed by centrifugation. Cell depleted supernatant is then filter sterilised through a 0.22μm filter. Filter sterilised supernatant is diluted 5 fold in 10mM sodium acetate pH 4.5. pH is adjusted by addition of 10 ml concentrated acetic acid per 5 liters of diluted supernatant. The ionic strength should be below 8mS/cm before application to the cation exchange column.

Diluted supernatant is loaded at a linear flow rate of 90cm/h onto a SP-sepharose FF (Pharmacia) column equilibrated with 50mM sodium acetate, pH 4.5 until the effluent from the column reaches a stable UV and conductivity baseline. To remove any unbound material, the column is washed using the equilibration buffer until the effluent from the column reaches a stable level with respect to UV absorbance and conductivity. The bound G-CSF protein is eluted from the column using a linear gradient; 30 column volumes; 0-80% buffer B (50mM NaAc, pH 4.5, 750mM NaCl) at a flow rate of 45cm/h. Based on SDS-polyacryl amide gel electrophoresis, fractions containing G-CSF are pooled. Sodium chloride is added until the ionic strength of the solution is more than 80mS/cm.

The protein solution is applied onto a Phenyl Toyo Pearl 650S column equilibrated with 50mM NaAc, pH 4.5, 750mM NaCl. Any unbound material is washed off the column using the equilibration buffer. Elution of G-CSF is performed by applying a step gradient of MilliQ water. Fractions containing G-CSF are pooled. By using this 2-step down stream processing strategy, more than 90% pure G-CSF can be obtained. The purified protein is then quantified using spectrophotometric measurements at 280nm and/or by amino acid analysis.

Fractions containing G-CSF are pooled. Buffer exchange and concentration is performed using VivaSpin concentrators (mwco: 5kDa).

EXAMPLE 4. IDENTIFICATION AND QUANTIFICATION OF NON-CONJUGATED AND CONJUGATED HG-CSF AND VARIANTS THEREOF

SDS-Polyacryl Amide Gel Electrophoresis

The purified, concentrated G-CSF was analyzed by SDS-PAGE. A single band having an apparent molecular weight of approx. 17kDa was dominant.

Absorbance

An estimate of the G-CSF concentration is obtained by spectrophotometric methods. By measuring the absorbance at 280 nm and using a theoretically extinction coefficient of 0.83, the protein concentration can be calculated.

Amino Acid Analysis

A more accurate protein determination can be obtained by amino acid analysis. Amino acid analysis performed on a purified G-CSF revealed that the experimentally determined amino acid composition is in agreement with the expected amino acid composition based on the DNA sequence.

EXAMPLE 5. MALDI-TOF MASS SPECTROMETRY OF PEGYLATED WT G-CSF AND G-CSF VARIANTS

MALDI-TOF mass spectrometry was used to evaluate the number of PEG-groups attached to PEGylated wt G-CSF and to selected PEGylated G-CSF variants.

Wt G-CSF contains 5 primary amines that are the expected attachment sites for SPA-PEG (the N-terminal amino-group and the ε-amino-group on K16, K23, K34 and K40). Following PEGylation of wt G-CSF with SPA-PEG 5000, MALDI-TOF mass spectrometry showed the presence of species of wt G-CSF with mainly 4, 5 and 6 PEG-groups attached. In addition, wt G-CSF with 7 PEG-groups attached was clearly seen although in minor amounts.

The G-CSF variant having the substitutions K16R, K34R, K40R, Q70K, Q90K, and Q120K also contains 5 primary amines (the N-terminal amino-group and the ε-amino-group on K23, K70, K90 and K120). Following PEGylation of this G-CSF variant with SPA-PEG5000, MALDI-TOF mass spectrometry showed the presence of species of the G-CSF variant with mainly 4, 5 and 6 PEG-groups attached. In addition, the G-CSF variant with 7 PEG-groups attached was clearly seen although in minor amounts.

The G-CSF variant having the substitutions K16R, K34R, and K40R contains 2 primary amines (the N-terminal amino-group and the ε-amino-group on K23). Following PEGylation of this G-CSF variant with SPA-PEG 12000, MALDI-TOF mass spectrometry showed the presence of species of the G-CSF variant with mainly 2 and 3 PEG-groups attached. In addition, the G-CSF variant with 4 PEG-groups attached was clearly seen although in minor amounts.

These observations clearly show that in addition to amino acid residues containing amine groups, other amino acid residues are sometimes PEGylated under the PEGylation conditions used. It also shows that it is of some importance for the PEGylation where amine groups are introduced. This has also been observed using SDS-PAGE analysis of wt G-CSF and G-CSF variants.

As described in Example 11, it has been shown that histidine 170 is fully PEGylated when the SPA-PEG chemistry is used. Furthermore, K23 and S159 are partly PEGylated. This explains the presence of 1-2 extra PEGylation sites besides the primary amines in hG-CSF and the variants that have been made.

EXAMPLE 6. PEPTIDE MAPPING OF PEGYLATED AND NON-PEGYLATED G-CSF VARIANTS

In order to map the additional attachment sites for SPA-PEG on G-CSF and G-CSF variants the following strategy was used.

A G-CSF variant with a low number of amine groups was chosen in order to reduce the number of expected PEGylation sites to a minimum. The G-CSF variant chosen has the substitutions K16R, K34R, K40R and H170Q. Apart from the ε-amino-group on K23 that previous data had shown not to be PEGylated to any large extent, this variant only contains one primary amine at the N-terminal. Thus, the background PEGylation on amine groups is significantly reduced in this G-CSF variant. The G-CSF variant was PEGylated using SPA-PEG 5000. Following PEGylation, the G-CSF variant was denatured, the disulphide bonds reduced, the resulting thiol groups alkylated, and the alkylated and PEGylated protein degraded with a glutamic acid-specific protease. Finally, the resulting peptides were separated by reversed phase HPLC.

Parallel with this, the non-PEGylated version of the G-CSF variant with the substitutions K16R, K34R, and K40R was treated identically in order to create a reference HPLC chromatogram.

Comparison of the HPLC chromatograms of the degradation of the PEGylated G-CSF variant and the non-PEGylated G-CSF variant should then reveal which peptides disappear upon PEGylation. Identification of these peptides by N-terminal amino acid sequencing of the peptide from the non-PEGylated G-CSF variant then indirectly points to the positions that are PEGylated.

In principle, it would have been preferable to use the non-PEGylated version of the G-CSF variant having all the substitutions K16R, K34R, K40R and H170Q, but for all practical purposes this does not matter.

More specifically, approximately 1 mg of the PEGylated G-CSF variant having the substitutions K16R, K34R, K40R and H170Q and approximately 500 µg of the non-PEGylated G-CSF variant having the substitutions K16R, K34R, and K40R were dried in a SpeedVac concentrator. The two samples were each dissolved in 400 µl 6 M guanidinium, 0.3 M Tris-HCl, pH 8.3 and denatured overnight at 37 °C. Following denaturation, the disulfide bonds in the proteins were reduced by addition of 50 µl 0.1 M DTT in 6 M guanidinium, 0.3 M Tris-HCl, pH 8.3. After 2 h of incubation at ambient temperature the thiol groups present were alkylated by addition of 50 µl 0.6 M iodoacetamid in 6 M guanidinium, 0.3 M Tris-HCl, pH 8.3. Alkylation took place for 30 min at ambient temperature before the reduced and alkylated proteins were buffer changed into 50 mM NH_4HCO_3 using NAP5 columns. The volumes of the samples were reduced to approximately 200 µl in a SpeedVac concentrator before addition of 20 µg and 10 µg glutamic acid-specific protease, respectively. The degradations with glutamic acid-specific protease were carried out for 16 h at 37 °C. The resulting peptides were separated by reversed phase HPLC employing a Phenomenex Jupiter C_{18} column (0.2 * 5 cm) eluted with a linear gradient of acetonitrile in 0.1% aqueous TFA. The collected fractions were analyzed by MALDI-TOF mass spectrometry and subsequently selected peptides were subjected to N-terminal amino acid sequence analysis.

Comparison of the HPLC chromatograms of the degradations of the PEGylated G-CSF variant and the non-PEGylated G-CSF variant revealed that only two fractions disappear upon PEGylation. N-terminal amino acid sequence analysis of the two fractions from the non-PEGylated G-CSF variant showed that the peptides both were derived from the N-terminal of G-CSF. One peptide consisted of amino acid residues 1-11 generated by an unexpected cleavage following Gln11. The other peptide consisted of amino acid residues 1-19 generated by an expected cleavage following Glu19.

It was expected that the N-terminal peptide of G-CSF would be identified using this approach, as the N-terminal amino group is easily PEGylated. However, none of the additional attachment sites for SPA-PEG 5000 were identified using this approach.

An alternative to the indirect identification of PEG 5000 attachment sites is direct identification of the attachment sites in PEGylated peptides. However, the fractions containing the PEGylated peptides in the HPLC separation of the degraded PEGylated G-CSF variant are poorly separated from each other and from several fractions containing non-PEGylated peptides. Thus, N-terminal amino acid sequence analysis of these fractions did not result in any useful data except for an indication that K23 could be partially PEGylated.

To overcome these problems, two pools of PEGylated peptides were made from the fractions from the first HPLC separation. These two pools were dried in a SpeedVac concentrator, dissolved in 200µl freshly prepared 50 mM NH_4HCO_3 and further degraded with 1 µg of chymotrypsin. The resulting peptides were separated by reversed phase HPLC employing a Phenomenex Jupiter C_{18} column (0.2 * 5 cm) eluted with a linear gradient of acetonitrile in 0.1% aqueous TFA. The collected fractions were analyzed by MALDI-TOF mass spectrometry and subsequently selected peptides were subjected to N-terminal amino acid sequence analysis.

From the N-terminal amino acid sequence determinations it could be determined that K23 as well as S159 are partially PEGylated. It was not possible to determine the exact degree of PEGylation at these two positions, but the PEGylation is only partial as peptides where K23 and S159 are unmodified were identified and sequenced from the initial HPLC separation.

EXAMPLE 7. GLYCOSYLATION OF WT G-CSF AND G-CSF VARIANTS

A consistent observation when analyzing purified wt G-CSF and G-CSF variants by MALDI-TOF mass spectrometry is the presence of an additional component with a mass approximately 324 Da larger than the mass of the G-CSF molecule analyzed. As the component with the lowest mass invariantly has the mass of the G-CSF molecule and because the G-CSF molecules have the correct N-terminal amino acid sequence, it was concluded that the additional component is a modified G-CSF molecule carrying two hexose residues. In many cases the unmodified G-CSF molecule gives rise to the most intense signal but in some cases the intensity of the signal for the modified G-CSF molecule is the most intense.

During the analysis of the peptides generated with the aim of identifying the additional PEGylation sites, two peptides of interest for identifying the site of glycosylation were identified in each of the degradations.

In both HPLC separations, the two peptides elute next to each other and MALDI-TOF mass spectrometry shows a mass difference between the two peptides of approximately 324 Da. The mass spectrometry data indicates that the peptide covers amino acid residues 124-162. N-terminal amino acid sequence analysis of all four peptides showed that this assignment is correct and that Thr133 is the only site of modification. In the peptides with the mass of the unmodified peptide, Thr133 is clearly seen in the sequence, while no amino acid residue can be assigned at position 133 in the peptides with an additional mass of 324 Da. As all other amino acid residues could be assigned in the sequence, it was concluded that Thr133 is the only site of modification. This glycosylation site was previously reported to be used in recombinant G-CSF expressed in CHO cells, although the glycan is different from the one attached by yeast.

The non-glycosylated wt G-CSF has been separated from the glycosylated wt G-CSF, employing reversed phase HPLC using a Vydac C_{18} column (0.21 * 5 cm) isocratically eluted with 51% acetonitrile in 0.1% TFA, as a fraction shown by MALDI-TOF mass spectrometry only to contain the non-glycosylated form of wt G-CSF.

EXAMPLE 8. SEPARATION OF G-CSF MOLECULES WITH DIFFERENT NUMBERS OF PEG MOLECULES COVALENTLY ATTACHED

Separation of G-CSF molecules covalently attached to 4, 5 or 6 PEG-groups was obtained as follows. PEGylated protein in 20mM sodium citrate, pH 2.5 was applied to an SP-sepharose FF column equilibrated with 20mM sodium citrate pH 2.5. Any unbound material was washed of the column. Elution was performed using a pH gradient. PEGylated G-CSF began to elute from the column at approx. pH 3.8 and continued to elute in fractions covering a pH span from 3.8 to 4.5.

The fractions were subjected to SDS-PAGE and mass spectrometric analysis. These analyses indicate that G-CSF having the highest degree of PEGylation is located in the "low pH fractions". PEGylated G-CSF having a lower degree of PEGylation is eluted in the "high pH fractions".

Amino acid analysis performed on PEGylated G-CSF showed good consistency between the theoretically and the experimentally determined extinction coefficient.

EXAMPLE 9. CONSTRUCTION OF HG-CSF VARIANTS

Specific substitutions of existing amino acids in hG-CSF to other amino acid residues, e.g., the specific substitutions discussed above in the general description, were introduced using standard DNA techniques known in the art. The new G-CSF variants were made using plasmid pG-CSFcerevisiae containing the gene, encoding hG-CSF without the HIS tag, as DNA template in the PCR reactions. The variants were expressed in *S. cerevisiae* and purified as described in Example 3. Some of the constructed G-CSF variants are listed below (see Example 9 and 10).

EXAMPLE 10. COVALENT ATTACHMENT OF SPA-PEG TO HG-CSF OR VARIANTS THEREOF

Human G-CSF and variants thereof were covalently linked to SPA-PEG 5000, SPA-PEG 12000 and SPA-PEG 20000 (Shearwater) as described above ("PEGylation of hG-CSF and variants thereof in solution"). The *in vitro* activities of the conjugates are listed in Example 12.

EXAMPLE 11. IDENTIFICATION OF SPA-PEG ATTACHMENT SITES IN G-CSF BY SITE-DIRECTED MUTAGENESIS FOLLOWED BY PEGYLATION OF THE PURIFIED VARIANTS

SPA-PEG may be attached to other amino acid residues than lysine in G-CSF. In order to determine whether SPA-PEG was attached to histidines, serines, threonines and arginines, variants were made in which these amino acids were substituted to lysine, alanine or glutamine. The variants were expressed in *S. cerevisiae*, purified and PEGylated followed by analysis of the number of attached SPA-PEG molecules on SDS-PAGE. A reduction in the number of attached SPA-PEG molecules after substitution of a given amino acid with glutamine or alanine strongly indicates that this amino acid is PEGylated by SPA-PEG, and this observation is further supported by an unchanged degree of PEGylation after substitution of the amino acid to lysine. The analyzed variants are listed below.

G-CSF variant	No. of attached PEG groups
hG-CSF	**10% 4 PEG, 75% 5 PEG, 15% 6 PEG**
K23R	10% 4 PEG, 85% 5 PEG, 5% 6 PEG
H43Q	10% 4 PEG, 75% 5 PEG, 15% 6 PEG
H43K	10% 5 PEG, 75% 6 PEG, 15% 7 PEG
H52Q	10% 4 PEG, 75% 5 PEG, 15% 6 PEG
H52K	10% 5 PEG, 75% 6 PEG, 15% 7 PEG

H156Q	10% 4 PEG, 75% 5 PEG, 15% 6 PEG
H156K	10% 5 PEG, 75% 6 PEG, 15% 7 PEG
H170Q	10% 3 PEG, 75% 4 PEG, 15% 5 PEG
H170K	10% 4 PEG, 75% 5 PEG, 15% 6 PEG
K16/34R	**10% 2 PEG, 75% 3 PEG, 15% 4 PEG**
K16/34R R22K	10% 3 PEG, 75% 4 PEG, 15% 5 PEG
K16/34R R22Q	10% 2 PEG, 75% 3 PEG, 15% 4 PEG
K16/34R S142A	10% 2 PEG, 75% 3 PEG, 15% 4 PEG
K16/34/40R	**10% 1 PEG, 75% 2 PEG, 15% 3 PEG**
K16/34/40R S53K	10% 2 PEG, 75% 3 PEG, 15% 4 PEG
K16/34/40R S53A	10% 2 PEG, 75% 3 PEG, 15% 4 PEG
K16/34/40R S62K	10% 2 PEG, 75% 3 PEG, 15% 4 PEG
K16/34/40R S66K	10% 2 PEG, 75% 3 PEG, 15% 4 PEG
K16/34/40R S80K	10% 2 PEG, 75% 3 PEG, 15% 4 PEG
K16/34/40R T105K	10% 2 PEG, 75% 3 PEG, 15% 4 PEG
K16/34/40R T133K	10% 2 PEG, 75% 3 PEG, 15% 4 PEG
K16/34/40R S142K	10% 2 PEG, 75% 3 PEG, 15% 4 PEG
K16/34/40R R147K	10% 2 PEG, 75% 3 PEG, 15% 4 PEG
K16/34/40R S155K	10% 2 PEG, 75% 3 PEG, 15% 4 PEG
K16/34/40R S159K	10% 2 PEG, 85% 3 PEG, 5% 4 PEG
K16/34/40R S170K	10% 1 PEG, 75% 2 PEG, 15% 3 PEG
K16/34/40R S170Q	5% 0 PEG, 85% 1 PEG, 15% 2 PEG

The data show that besides the N-terminus, K16, K34 and K40, SPA-PEG also is covalently bound to H170. Furthermore, the data show that only 10% of the available K23 amino acid residues are PEGylated, and that approximately 10% of S159 is PEGylated.

EXAMPLE 12. IN VITRO BIOLOGICAL ACTIVITY OF NON-CONJUGATED AND CONJUGATED HG-CSF AND VARIANTS THEREOF

The *in vitro* biological activities of conjugated and non-conjugated hG-CSF and variants thereof were measured as described above in "Primary assay 2 – *in vitro* hG-CSF activity assay". The measured EC50 values for each variant with and without conjugation of SPA-PEG 5000 to the available PEGylation sites are listed below. The values have been normalized with respect to the EC50 value of non-conjugated hG-CSF (Neupogen®). This value was measured simultaneously with the variants each time under identical assay conditions. The EC50 value of hG-CSF in the described assay is 30 pM.

G-CSF variant	EC50 (% of hG-CSF) non-conjugated	EC50 (% of hG-CSF) conjugated to SPA-PEG 5000
G-CSF with N-terminal Histidine tag	10	Not determined
G-CSF without N-terminal Histidine tag	100	0.1
16R	100	1
16Q	80	1
23Q	80	0.1
23R	100	0.1
34R	100	1
34A	80	1
34Q	70	1
40R	50	1
K16/23R	100	1
K16/23Q	80	1
K34/40R	50	5
K16/34R	100	10
K16/40R	50	5
K16/23/34R	50	10
K16/23/40R	50	5
K16/34/40R	**35**	**30**
K16/23/34/40R	20	15
K16/34/40R E45K	Expressed at low levels	Not determined
K16/34/40R E46K	10	1
K16/34/40R S53K	5	0.5
K16/34/40R S62K	10	0.5
K16/34/40R S66K	20	2
K16/34/40R Q67K	10	0.2
K16/34/40R Q70K	**30**	**20**
K16/34/40R S80K	10	0.2
K16/34/40RQ90K	**30**	**20**
K16/34/40R E98K	Expressed at low levels	Not determined
K16/34/40R D104K	10	0.9
K16/34/40R T105K	**30**	**10**
K16/34/40R Q120K	**30**	**20**
K16/34/40R Q131K	Expressed at low levels	Not determined
K16/34/40R T133K	**30**	**10**
K16/34/40R Q134K	10	0.2
K16/34/40R S142K	20	7
K16/34/40R R147K	20	1
K16/34/40R S155K	20	1
K16/34/40R S159K	20	3

K16/34/40R Q70K Q120K	25	15
K16/34/40R Q90K Q120K	25	15
K16/34/40R T105K Q120K	20	8
K16/34/40R Q120K T133K	20	8
K16/34/40R Q120K S142K	10	2
K16/34/40R Q70K Q90K T105K	10	2
K16/34/40R Q70K Q90K Q120K	20	10
K16/34/40R Q70K Q90K T133K	15	5
K16/34/40R Q70K T105K Q120K	10	2
K16/34/40R Q70K Q120K T133K	15	5
K16/34/40R Q70K Q120K S142K	10	1
K16/34/40R Q90K T105K Q120K	10	2
K16/34/40R Q90K T105K T133K	10	2
K16/34/40R Q90K Q120K T133K	15	5
K16/34/40R Q90K Q120K S142K	10	1
K16/34/40R T105K Q120K T133K	10	1
K16/34/40R Q120K T133K S142K	10	1

The data show that substitution of K23 to arginine does not increase the activity of the conjugated protein. This is due to the fact that only 10% of K23 is PEGylated, whereby the conjugated K23R variant has essentially the same number of PEG groups attached to it and has the same location of the PEGylation sites as hG-CSF. Removal of the remaining lysines at position K16, K34 and K40 resulted in a G-CSF variant with significant activity after PEGylation. Conjugation of SPA-PEG 5000 to this variant does not decrease the activity significantly as compared to the non-conjugated variant. Thus, PEGylation of the N-terminus and H170 with SPA-PEG 5000 does not decrease the activity of hG-CSF. It was decided to use hG-CSF K16R K34R K40R as the backbone for insertion of new PEGylation sites. 21 new PEGylation sites between residues E45 and H159 were introduced in this backbone. These residues are distributed over the parts of hG-CSF that do not interact with the G-CSF receptor. Introduction of new PEGylation sites at positions Q70, Q90, T105, Q120, T133 and S142 resulted in hG-CSF variants that retained a significant amount of activity after PEGylation by SPA-PEG 5000. Thus, these new PEGylation sites were combined in hG-CSF variants that had 2 or 3 new PEGylation sites. The most active of these variants after PEGylation with SPA-PEG 5000 were hG-CSF K16R K34R K40R Q70K Q120K and hG-CSF K16R K34R K40R Q70K Q90K Q120K.

Furthermore, SPA-PEG 5000, SPA-PEG 12000 and SPA-PEG 20000 were attached to hG-CSF K16R K34R K40R. The *in vitro* EC50 values were 35%, 10% and 1%, respectively, of non-conjugated hG-CSF (Neupogen®).

EXAMPLE 13. IN VIVO HALF-LIFE OF NON-CONJUGATED AND CONJUGATED HG-CSF AND VARIANTS THEREOF

The *in vivo* half-lives of non-conjugated hG-CSF (Neupogen®) and SPA-PEG 5000 conjugated hG-CSF K16R K34R K40R Q70K Q90K Q120K were measured as described above ("Measurement of the *in vivo* half-life of conjugated and non-conjugated rhG-CSF and variants thereof"). The results are shown in Figure 1. The *in vivo* half-life of Neupogen® and SPA-PEG 5000 conjugated hG-CSF K16R K34R K40R Q70K Q90K Q120K was determined to be 2.01 hours and 12.15 hours, respectively. Thus, introducing new PEGylation sites in hG-CSF and conjugating SPA-PEG 5000 to them has resulted in a significant increase in the *in vivo* half-life. In an earlier experiment, the *in vivo* half-life of Neupogen® was compared to hG-CSF with a larger N-terminally conjugated PEG molecule (10 kDa). Here, the *in vivo* half-lives were determined to be 1.75 hours and 7.01 hours, respectively. Thus, SPA-PEG 5000 conjugated hG-CSF K16R K34R K40R Q70K Q90K Q120K has a significantly longer half-life than both Neupogen® and hG-CSF with a 10 kDa N-terminally conjugated PEG molecule.

EXAMPLE 14. IN VIVO BIOLOGICAL ACTIVITY OF NON-CONJUGATED AND CONJUGATED HG-CSF AND VARIANTS THEREOF

The *in vivo* biological activities of non-conjugated hG-CSF (Neupogen®), SPA-PEG 5000 conjugated hG-CSF K16R K34R K40R Q70K Q120K and SPA-PEG 5000 conjugated hG-CSF K16R K34R K40R Q70K Q90K Q120K were measured as described above ("Measurement of the *in vivo* biological activity of conjugated and non-conjugated hG-CSF and variants thereof"). The results are shown in Figure 2. No activity of Neupogen® could be detected at 48 hours after injection of 100 μg per kg body weight at t=0 hours. Activity of SPA-PEG 5000 conjugated hG-CSF K16R K34R K40R Q70K Q120K could be detected until 72 hours after the initial injection, while SPA-PEG 5000 conjugated hG-CSF K16R K34R K40R Q70K Q90K Q120K remained active *in vivo* until 96 hours after the initial injection. Thus, it was shown that both conjugated variants had a longer *in vivo* biological activity than Neupogen® and

that SPA-PEG 5000 conjugated hG-CSF K16R K34R K40R Q70K Q90K Q120K remained active twice as long *in vivo* as Neupogen®.

Furthermore, the *in vivo* biological activities of Neupogen®, SPA-PEG 12000 conjugated hG-CSF K16R K34R K40R and different doses of SPA-PEG 5000 conjugated hG-CSF K16R K34R K40R Q70K Q90K Q120K were measured. The results are shown in Figure 3. As observed earlier, no activity of Neupogen® could be detected 48 hours after the initial injection of 100 μg per kg body weight. Administration of 5 μg per kg body weight of SPA-PEG 5000 conjugated hG-CSF K16R K34R K40R Q70K Q90K Q120K resulted in a slightly longer *in vivo* biological activity than Neupogen®, while administration of 25 μg per kg body weight and 100 μg per kg body weight of this compound resulted in hG-CSF activity until 72 and 96 hours, respectively, after the initial injection. SPA-PEG 12000 conjugated hG-CSF K16R K34R K40R remained active *in vivo* until 72 hours after administration of 100 μg per kg body weight. As described in Example 5, SPA-PEG 12000 conjugated hG-CSF K16R K34R K40R has 2 or 3 SPA-PEG 12000 groups attached while SPA-PEG 5000 conjugated hG-CSF K16R K34R K40R Q70K Q90K Q120K has 5 or 6 SPA-PEG 5000 groups attached. Thus, the molecular weights of the two compounds are 42-54 kDa and 43-48 kDa, respectively. The longer *in vivo* biological activity of SPA-PEG 5000 conjugated hG-CSF K16R K34R K40R Q70K Q90K Q120K as compared to SPA-PEG 12000 conjugated hG-CSF K16R K34R K40R with essentially the same molecular weight shows that the distribution of the attached PEG molecules is important for the duration of the *in vivo* biological activity.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques, methods, compositions, apparatus and systems described above may be used in various combinations. All publications, patents, patent applications, or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, or other document were individually indicated to be incorporated by reference for all purposes.

SEQUENCE LISTING

The appended sequence listing contains the following sequences:

SEQ ID NO:1: The amino acid sequence of human G-CSF.

SEQ ID NO:2: A synthetic DNA sequence encoding human G-CSF, with codon usage optimised for expression in *E. coli*.

SEQ ID NO:3: The amino acid sequence of the OmpA signal sequence.

SEQ ID NO:4: A synthetic DNA sequence encoding the OmpA signal sequence.

SEQ ID NO:5: A synthetic histidine tag.

SEQ ID NO:6: A synthetic DNA sequence encoding the histidine tag of SEQ ID NO:5.

SEQ ID NO:7: The amino acid sequence of a human G-CSF signal peptide.

SEQ ID NO:8: A synthetic DNA sequence encoding human G-CSF, including the signal peptide of SEQ ID NO:7, with codon usage optimised for expression in CHO cells.

<110> Maxygen ApS

<120> G-CSF conjugates

<130> 8us02

<150> DK PA 2000 00024
<151> 2000-01-10

<150> US 60/176,376
<151> 2000-01-14

<150> DK PA 2000 00341
<151> 2000-03-02

<150> US 60/189,506
<151> 2000-03-15

<150> DK PA 2000 00943
<151> 2000-06-16

<150> US 60/215,644
<151> 2000-06-30

<160> 8

<170> PatentIn version 3.0

<210> 1
<211> 174
<212> PRT
<213> Homo sapiens

<400> 1

```
Thr Pro Leu Gly Pro Ala Ser Ser Leu Pro Gln Ser Phe Leu Leu Lys
1               5                  10                  15

Cys Leu Glu Gln Val Arg Lys Ile Gln Gly Asp Gly Ala Ala Leu Gln
            20                  25                  30

Glu Lys Leu Cys Ala Thr Tyr Lys Leu Cys His Pro Glu Glu Leu Val
        35                  40                  45

Leu Leu Gly His Ser Leu Gly Ile Pro Trp Ala Pro Leu Ser Ser Cys
    50                  55                  60

Pro Ser Gln Ala Leu Gln Leu Ala Gly Cys Leu Ser Gln Leu His Ser
65                  70                  75                  80

Gly Leu Phe Leu Tyr Gln Gly Leu Leu Gln Ala Leu Glu Gly Ile Ser
                85                  90                  95

Pro Glu Leu Gly Pro Thr Leu Asp Thr Leu Gln Leu Asp Val Ala Asp
            100                 105                 110
```

Phe Ala Thr Thr Ile Trp Gln Gln Met Glu Glu Leu Gly Met Ala Pro
115 120 125

Ala Leu Gln Pro Thr Gln Gly Ala Met Pro Ala Phe Ala Ser Ala Phe
130 135 140

Gln Arg Arg Ala Gly Gly Val Leu Val Ala Ser His Leu Gln Ser Phe
145 150 155 160

Leu Glu Val Ser Tyr Arg Val Leu Arg His Leu Ala Gln Pro
165 170

<210> 2
<211> 525
<212> DNA
<213> Artificial sequence

<400> 2
acccctctgg gcccggccag cagtctgcct cagagttttt tactgaaatg cttagaacag 60

gtgcgtaaaa tccagggcga tggcgcggcc ctgcaggaaa aactgtgcgc gacctataaa 120

ctgtgccatc ctgaagaact ggtcctgtta ggccatagct taggcatccc gtgggcgcct 180

ctgagtagct gcccgagtca ggccctgcag ctggccggct gcctgagtca gttacatagt 240

ggcttatttt tatatcaggg cttactgcag gcgttagaag gcattagtcc ggaactgggc 300

ccgaccctgg ataccttaca gttagatgtc gcggattttg ccaccaccat ttggcagcag 360

atggaagaat taggcatggc gcctgcgtta cagcctaccc agggcgccat gcctgcgttt 420

gcgagtgcgt ttcagcgtcg cgccggcggc gtgttagtgg ccagccatct gcagagcttt 480

ctggaagtga gttatcgtgt gttacgccat ctggcccagc cttaa 525

<210> 3
<211> 21
<212> PRT
<213> Escherichia coli

<400> 3

Met Lys Lys Thr Ala Ile Ala Ile Ala Val Ala Leu Ala Gly Phe Ala
1 5 10 15

Thr Val Ala Gln Ala
20

<210> 4
<211> 63
<212> DNA
<213> Artificial sequence

<400> 4

atgaaaaaga cagctatcgc gattgcagtg gcactggctg gtttcgctac cgtagcgcag 60

gcc 63

<210> 5
<211> 15
<212> PRT
<213> Artificial sequence

<400> 5

Met Lys His Gln His Gln His Gln His Gln His Gln His Gln Gln
1 5 10 15

<210> 6
<211> 45
<212> DNA
<213> Artificial sequence

<400> 6
atgaaacacc aacaccaaca tcaacatcaa catcaacatc aacag 45

<210> 7
<211> 30
<212> PRT
<213> Homo sapiens

<400> 7

Met Ala Gly Pro Ala Thr Gln Ser Pro Met Lys Leu Met Ala Leu Gln
1 5 10 15

Leu Leu Leu Trp His Ser Ala Leu Trp Thr Val Gln Glu Ala
20 25 30

<210> 8
<211> 615
<212> DNA
<213> Artificial sequence

<400> 8
atggccggcc ctgccacaca gtcccccatg aagctgatgg ccctgcagct gctgctgtgg 60

cactccgccc tgtggacagt gcaggaggcc acccctctgg gccccgccag ctccctgcct 120

cagtccttcc tgctgaagtg cctggagcag gtgagaaaga tccagggcga cggcgccgcc 180

ctgcaggaga agctgtgcgc cacatacaag ctgtgccacc ctgaggagct ggtgctgctg 240

ggccacagcc tgggcatccc ctgggccccc ctgtccagct gcccctccca ggccctgcag 300

ctggccggct gcctgtccca gctgcactcc ggcctgttcc tgtaccaggg cctgctgcag 360

gccctggagg gcatctcccc cgagctgggc cccacactgg ataccctgca gctggacgtg 420

gccgatttcg ccaccacaat ctggcagcag atggaggagc tgggcatggc ccctgccctg 480

cagcctaccc agggcgccat gcctgccttt gcctccgcct ttcagagacg ggccggcggc 540

gtgctggtgg ccagccacct gcagagcttt ctggaggtgt cctacagagt gctgcggcac 600

ctggcccagc cttga 615